

Lower bounds for CSP refutation by SDP hierarchies

Ryuhei Mori*

David Witmer†

October 11, 2016

Abstract

For a k -ary predicate P , a random instance of $\text{CSP}(P)$ with n variables and m constraints is unsatisfiable with high probability when $m \gg n$. The natural algorithmic task in this regime is *refutation*: finding a proof that a given random instance is unsatisfiable. Recent work of Allen et al. suggests that the difficulty of refuting $\text{CSP}(P)$ using an SDP is determined by a parameter $\text{cmplx}(P)$, the smallest t for which there does not exist a t -wise uniform distribution over satisfying assignments to P . In particular they show that random instances of $\text{CSP}(P)$ with $m \gg n^{\text{cmplx}(P)/2}$ can be refuted efficiently using an SDP.

In this work, we give evidence that $n^{\text{cmplx}(P)/2}$ constraints are also *necessary* for refutation using SDPs. Specifically, we show that if P supports a $(t - 1)$ -wise uniform distribution over satisfying assignments, then the Sherali-Adams₊ and Lovász-Schrijver₊ SDP hierarchies cannot refute a random instance of $\text{CSP}(P)$ in polynomial time for any $m \leq n^{t/2-\varepsilon}$.

*Department of Mathematical and Computing Sciences, Tokyo Institute of Technology. mori@is.titech.ac.jp. Supported by MEXT KAKENHI Grant Number 24106008.

†Computer Science Department, Carnegie Mellon University. dwitmer@cs.cmu.edu. Supported by the National Science Foundation Graduate Research Fellowship Program under grant DGE-1252522, by NSF grants CCF-0747250 and CCF-1116594, and by a CMU Presidential Fellowship.

1 Introduction

The average-case complexity of constraint satisfaction problems (CSPs) has been studied extensively in computer science, mathematics, and statistical physics. Despite the vast amount of research that has been done, the hardness of natural algorithmic tasks for random CSPs remains poorly understood. Given a k -ary predicate $P : \{0, 1\}^k \rightarrow \{0, 1\}$, we consider random instances of $\text{CSP}(P)$ with n variables and m constraints. Each constraint is chosen independently and consists of P applied to k literals (variable or their negations) chosen independently and uniformly at random. Whether or not a random CSP is satisfiable depends on its clause density $\frac{m}{n}$. It is conjectured that for any nontrivial CSP there is a *satisfiability threshold* $\alpha(P)$ depending on the choice of predicate P : For $m < \alpha(P) \cdot n$, an instance is satisfiable with high probability, and $m > \alpha(P) \cdot n$, an instance is unsatisfiable with high probability. This conjecture has been proven in the case of k -SAT for large enough k [DSS15], but, to our knowledge, remains open for all other predicates. Even so, it is easy to show that when $m \gg n$, an instance is unsatisfiable with high probability. In the low density, satisfiable regime, the major research goal is to develop algorithms that find satisfying assignments. In the high density, unsatisfiable regime, the goal is to *refute* an instance, i.e., find a short certificate that there is no solution.

In this paper, we study refutation. A refutation algorithm takes a random instance \mathcal{I} of $\text{CSP}(P)$ and returns either “unsatisfiable” or “don’t know”. It must satisfy two conditions: (1) it is never wrong, i.e., if \mathcal{I} is satisfiable, it must return “don’t know” and (2) it returns “unsatisfiable” with high probability over the choice of the instance. As m increases, refutation becomes easier. The objective, then, is to refute instances with m as small as possible. This problem has been studied extensively and is related to hardness of approximation [Fei02], proof complexity [BSB02], statistical physics [CLP02], cryptography [ABW10], and learning theory [DLSS14]. Much research has focused on finding algorithms for refutation, especially in the special case of SAT; see [AOW15] for references.

Most known refutation algorithms are based on semidefinite programming (SDP). For now, we think of an SDP relaxation of an instance \mathcal{I} of $\text{CSP}(P)$ as a black box that returns a number $\text{SDPOpt} \in [0, 1]$ that upper bounds the maximum fraction of constraints that can be simultaneously satisfied. An SDP-based refutation algorithm takes a random instance \mathcal{I} of $\text{CSP}(P)$, solves some SDP relaxation of \mathcal{I} , and return “Unsatisfiable” if and only if $\text{SDPOpt} < 1$. Many known polynomial-time algorithms for refutation fit into this framework (e.g., [AOW15, BM15, FGK05, COGL04, FO04]). It is then natural to ask the following question.

What is the minimum number of constraints needed to refute random instances of $\text{CSP}(P)$ using an efficient SDP-based refutation algorithm?

Allen et al. give an upper bound on the number of constraints required to refute an instance of $\text{CSP}(P)$ in terms of a parameter $\text{cmplx}(P)$, defined to be the minimum t such that there is no t -wise uniform distribution over satisfying assignments to P [AOW15].

Theorem 1.1 ([AOW15]). *There is an efficient SDP-based algorithm that refutes a random instance \mathcal{I} of $\text{CSP}(P)$ with high probability when $m \gg n^{\text{cmplx}(P)/2}$.*

Clearly, $1 \leq \text{cmplx}(P) \leq k$ for nontrivial predicates. Also, $\text{cmplx}(P) = k$ when P is k -XOR or k -SAT.

For special classes of predicates, we know that $n^{\text{cmplx}(P)/2}$ constraints are also necessary for refutation by SDP-based algorithms. Schoenebeck considered arity- k predicates P whose satisfying assignments are a superset of k -XOR’s; these include k -SAT and k -XOR. For such predicates, he showed that polynomial-size sum of squares (SOS) SDP relaxations cannot refute random instances with $m \leq n^{k/2-\varepsilon}$ [Sch08] using a proof previously discovered by Grigoriev [Gri01]. Lee, Raghavendra, and Steurer showed that the SOS relaxation of $\text{CSP}(P)$ is at least as powerful as an arbitrary SDP relaxation of comparable size [LRS15]. With Schoenebeck’s result, this implies that no polynomial-size SDP can be used to refute random instances of k -XOR or k -SAT when $m \leq n^{k/2-\varepsilon}$. This leads us to make the following conjecture.

Conjecture 1.2. *Let ε be a constant greater than 0. Given a random instance \mathcal{I} of $\text{CSP}(P)$ with $m \leq n^{\text{cmplx}(P)/2-\varepsilon}$, with high probability any polynomial-size SDP relaxation of \mathcal{I} has optimal value 1 and can therefore not be used to refute \mathcal{I} .*

It suffices to prove this conjecture for SOS SDP relaxations [LRS15]. This would essentially complete our understanding of the power of SDP-based refutation algorithms. Prior to this work, this SOS version Conjecture 1.2 appeared in [AOW15]; we know of no other mention of this conjecture in the literature.¹

Note that in the special case of k -XOR, SDP algorithms are *not* optimal. A k -XOR constraint is a linear equation mod 2, so Gaussian elimination can be used to refute *any* unsatisfiable k -XOR instance. A random k -XOR instance with $m \gg n$ can therefore be refuted with high probability. SDP-based algorithms, on the other hand, require $m \geq n^{k/2}$ [Gri01, Sch08, LRS15]. More generally, if P can be written as a degree- d polynomial over \mathbb{F}_2 , then Gaussian elimination can be used to refute random instances with $m = O(n^d)$. See [OW14] for more details.

Some partial progress has been made toward proving Conjecture 1.2. Building on results of Benabbas et al. [BGMT12] and Tulsiani and Worah [TW13], O’Donnell and Witmer proved lower bounds for the Sherali-Adams (SA) linear programming (LP) hierarchy and the Sherali-Adams₊ (SA₊) and Lovász-Schrijver₊ (LS₊) SDP hierarchies. All three of these hierarchies are weaker than SOS. The SA₊ hierarchy gives an optimal approximation for any CSP in the worst case assuming the Unique Games Conjecture [Rag08]. They showed that polynomial-size Sherali-Adams linear programming (LP) relaxations cannot refute random instances with $m \leq n^{\text{cmplx}(P)/2-\varepsilon}$ [OW14]; this implies that no polynomial-size LP can refute random instances with $m \leq n^{\text{cmplx}(P)/2-\varepsilon}$ by work of Chan et al. [CLRS13]. They also showed that SA₊ cannot refute random instances with $m \leq n^{\text{cmplx}(P)/2-\varepsilon}$ when a set of $o(m)$ constraints has been removed [OW14]. Also, they proved that SA₊ and LS₊ cannot refute fully random instances with $m \leq n^{\text{cmplx}(P)/2-1/3-\varepsilon}$. Much less is known for SOS. For predicates P that support a pairwise uniform distribution over satisfying assignments, Barak, Chan, and Kothari showed that polynomial-size SOS relaxations cannot refute random instances with $m = \Omega(n)$ in which $o(m)$ constraints have been removed [BCK15].

In addition, there is a long history of related work on lower bounds for refutation in proof complexity (e.g., [Gri01, KI06, BOGH⁺06, TW13]). Specifically, SA, SA₊, LS₊, and SOS have corresponding static semialgebraic proof systems and proving integrality gaps for these LP and SDP relaxations is equivalent to proving rank or degree lower bounds for refutations in these proof systems.

Results Our contribution is two-fold: First, we remove the assumption that a small number of constraints are deleted to show that fully-random CSP instances have integrality gaps in SA₊ for $m \leq n^{t/2-\varepsilon}$. As in [BGMT12] and [OW14], we prove this result for predicates over possibly larger alphabets $[q]$.

Theorem 1.3. *Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t-1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability the SA₊ relaxation for \mathcal{I} has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

Second, we use this result to show that fully random instances have LS₊ integrality gaps for $m \leq n^{t/2-\varepsilon}$. Recall that LS₊ gives relaxations of 0/1-valued integer programs, so we restrict our attention here to Boolean CSPs with $P : \{0, 1\}^k \rightarrow \{0, 1\}$.

Theorem 1.4. *Let $P : \{0, 1\}^k \rightarrow \{0, 1\}$ be $(t-1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability the LS₊ relaxation for \mathcal{I} has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

¹Barak, Kindler, and Steurer [BKS13] made a related but different conjecture that the basic SDP relaxation is optimal for random CSPs.

In their strongest form, our results hold for a static variant of the LS_+ SDP hierarchy that is at least as strong as both SA_+ and LS_+ and is dual to the static LS_+ proof system studied in previous work (e.g., [GHP02, KI06]). We define this static LS_+ hierarchy in Section 2.

Theorem 1.5. *Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability the static LS_+ relaxation for \mathcal{I} has value 1, even after $\Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.*

Tulsiani and Worah proved this theorem in the special case of pairwise independence and $O(n)$ constraints [TW13, Theorem 3.27].² These results provide further evidence for Conjecture 1.2 and, in particular, give the first examples of SDP hierarchies that are unable to refute CSPs with $(t - 1)$ -wise uniform-supporting predicates when $m \leq n^{t/2-\varepsilon}$.

From a dual point of view, we can think of SA_+ , LS_+ , and static LS_+ as semialgebraic proof systems and our results can be equivalently stated as rank or degree lower bounds for these proof systems.

Theorem 1.6. *Let $P : \{0, 1\}^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability any SA_+ or static LS_+ refutation of \mathcal{I} requires degree $\Omega(n^{\frac{\varepsilon}{t-2}})$ and any LS_+ refutation of \mathcal{I} requires rank $\Omega(n^{\frac{\varepsilon}{t-2}})$.*

In another line of work, Feldman, Perkins, and Vempala [FPV15] showed that if a predicate P is $(t - 1)$ -wise uniform supporting, then any statistical algorithm based on an oracle taking L values requires $m = \tilde{\Omega}(n^t/L)$ to refute. They further show that the dimension of any convex program refuting such a CSP must be at least $\tilde{\Omega}(n^{t/2})$. These lower bounds are incomparable to the integrality gap results stated above: While statistical algorithms and arbitrary convex relaxations are more general computational models, standard SDP hierarchy relaxations for k -CSPs, including the SA_+ and LS_+ relaxations we study, have dimension $n^{O(k)}$ and are therefore not ruled out by this work.

Techniques To solve $\text{CSP}(P)$ exactly, it suffices to optimize the expected fraction of satisfied constraints over distributions on assignments $[q]^n$. This, of course, is hard, so relaxations like SA , SA_+ , LS_+ , and SOS instead optimize over local distributions on assignments to smaller sets of variables. As the number of rounds of the relaxation increases, we look at distributions over assignments to larger and larger sets of variables; the r -round relaxation considers distributions over assignments to sets of size at most r and has size $n^{O(r)}$. The r -round SA_+ relaxation requires that (1) local distributions on assignments to sets of at most r variables satisfy consistency conditions and (2) the covariance matrix corresponding to these local distributions is positive semidefinite (PSD); see Section 2 for precise definitions. We know that when $m \leq n^{\text{cmplx}(P)/2-\varepsilon}$, there exist local distributions satisfying (1) [BGMT12, OW14].

When the number of constraints is $O(n)$, previous work showed that the covariance matrix corresponding to the [BGMT12] local distributions is PSD by proving that it is diagonal and has nonnegative entries. If the covariance matrix is diagonal, then there is no correlation between assignments to pairs of variables under the corresponding local distributions. This condition holds for instances with number of constraints small enough, but correlations between variables arise as the number of constraints increases.

We prove PSDness in the presence of these correlations by showing that they must remain local. Our argument extends a technique of Tulsiani and Worah [TW13]. We prove that the graph induced by correlations between variables must have small connected components, each of which corresponds to a small block of nonzero entries in the covariance matrix. Since these blocks are small, condition (1) guarantees that for each block there exists an actual distribution on assignments to the variables of that block. This means that each of these blocks is the covariance matrix of an actual distribution and must therefore be PSD. The entire covariance matrix must then be PSD: It can be written as a block diagonal matrix in which each block is PSD.

²Actually, [TW13] prove a rank lower bound for the dual static LS_+ proof system, but this is equivalent to a rank lower bound for the static LS_+ SDP hierarchy we consider here.

2 Preliminaries

2.1 Constraint satisfaction problems

Definition 2.1. Given a predicate $P : [q]^k \rightarrow \{0, 1\}$, an instance \mathcal{I} of the $\text{CSP}(P)$ problem with variables x_1, \dots, x_n is a multiset of P -constraints. Each P -constraint is a tuple (c, S) , where $c \in [q]^k$ is the negation pattern and $S \in [n]^k$ is the scope. The corresponding constraint is $P(x_S + c) = 1$, where $x_S = (x_i)_{i \in S}$ and $+$ is component-wise addition mod q .

In the decision version of $\text{CSP}(P)$, we want to determine whether or not there exists an assignment satisfying all constraints of a given instance \mathcal{I} . In the optimization version, the objective is to maximize the fraction of simultaneously satisfied constraints. That is, we define $\text{Val}_{\mathcal{I}}(x) = \frac{1}{m} \sum_{(c, S) \in \mathcal{I}} P(c + x_S)$ and want to find $x \in [q]^n$ maximizing $\text{Val}_{\mathcal{I}}(x)$. We will write $\max_x \text{Val}_{\mathcal{I}}(x)$ as $\text{Opt}(\mathcal{I})$.

We will show that SOS cannot solve the decision version for random instances \mathcal{I} with small enough number of constraints even though such instances are far from satisfiable. This implies that SOS cannot show that $\text{Opt}(\mathcal{I}) < 1$ for such instances.

Next, we define our random model. We consider instances in which m constraints are drawn independently and uniformly at random from among all $q^k n^k$ possible constraints with replacement. We distinguish between different orderings of the scope, as P may not be symmetric. The specific details of this definition are not important; our results hold for any similar model. For example, see [AOW15, Appendix D]. A random instance is likely to be highly unsatisfiable: It is easy to show that $\text{Opt}(\mathcal{I}) = \frac{|P^{-1}(1)|}{q^k} + o(1)$ for $m \geq n \log n$ with high probability.

Given an instance \mathcal{I} , the associated k -uniform hypergraph $H_{\mathcal{I}}$ on $V = [n]$ has a hyperedge S if and only if S is the scope of some constraint of \mathcal{I} . Here, we disregard the orderings of the constraint scopes. Given a hypergraph H and a subset of vertices T , we let $H[T]$ be the subhypergraph induced by T .

We consider predicates for which there exist distributions over satisfying assignments that look uniform on every small enough set of bits. Formally, we study the following condition.

Definition 2.2. A predicate $P : [q]^k \rightarrow \{0, 1\}$ is t -wise uniform supporting if there exists a distribution μ over $[q]^k$ supported on P 's satisfying assignments such that $\Pr_{z \sim \mu}[z_T = \alpha] = q^{-|T|}$ for all $\alpha \in [q]^{|T|}$ and for all $T \subseteq [k]$ with $1 \leq |T| \leq t$.

2.2 LP and SDP hierarchies

2.2.1 Representing $\text{CSP}(P)$ with polynomial inequalities

An LP or SDP hierarchy is a procedure for constructing increasingly tight relaxations of a set of polynomial inequalities. In our case, these polynomial inequalities represent the constraints of a CSP. In this section, we describe two standard ways of writing these polynomial formulations. Both encode the decision version of $\text{CSP}(P)$. The relaxations based on these encodings are at least as strong as corresponding relaxations of the maximization version of $\text{CSP}(P)$, so our lower bounds for the decision problem imply lower bounds for the maximization problem. We describe the case of binary alphabets first and then mention how CSPs with larger alphabets can be encoded using binary variables.

For SA, SA₊, and static LS₊ relaxations, we represent each constraint as a degree- k polynomial equality. Let $P'(x)$ be the unique multilinear degree- k polynomial such that $P'(z) = P(z)$ for all $z \in \{0, 1\}^k$. Also, given $a \in [0, 1]^k$ and $b \in \{0, 1\}$, use $a^{(b)}$ to denote a if $b = 0$ and $1 - a$ if $b = 1$. For $z \in [0, 1]^k$ and $c \in \{0, 1\}^k$, let $z^{(c)} \in [0, 1]^k$ be such that $(z^{(c)})_i = z_i^{(c_i)}$. The degree- k formulation of \mathcal{I} is defined as follows.

$$R_{\mathcal{I}} = \{P'(x_S^{(c)}) = 1 \mid (c, S) \in \mathcal{I}\}. \quad (2.1)$$

For LS_+ , on the other hand, we have to start with a set of *linear* inequalities and only consider the binary alphabet case. Recall that any nontrivial arity- k predicate P can be represented as a conjunction of at most $2^k - 1$ disjunctions of arity k . In particular, letting $F = \{z \in \{0, 1\}^k \mid P(z) = 0\}$, we see that

$$P(z) = \bigwedge_{f \in F} \bigvee_{i=1}^k f_i \oplus z_i. \quad (2.2)$$

Using (2.2), we can represent \mathcal{I} as a k -SAT instance with at most $(2^k - 1) \cdot m$ constraints and then consider the standard linear relaxation of this k -SAT instance. For each clause $\bigvee_{i=1}^k c_i \oplus z_i$, we add the inequality $\sum_{i=1}^k z_i^{(c_i)} \geq 1$ and obtain the following linear relaxation for \mathcal{I} .

$$L_{\mathcal{I}} = \left\{ \sum_{i=1}^k x_{S_i}^{(c_i \oplus f_i)} \geq 1 \mid (c, S) \in \mathcal{I}, f \in F \right\}. \quad (2.3)$$

It is more natural to apply SA, SA_+ , and static LS_+ to (2.1), but applying SA, SA_+ , and static LS_+ to (2.3) yields relaxations that are approximately equivalent.

Lemma 2.3. *Let $r \geq k$ and let \mathcal{I} be an instance of $\text{CSP}(P)$ with binary alphabet. Then the following statements hold.*

1. $\text{SA}^r(R_{\mathcal{I}}) \subseteq \text{SA}^{r+k+1}(L_{\mathcal{I}})$ and $\text{SA}^r(L_{\mathcal{I}}) \subseteq \text{SA}^{r+k+1}(R_{\mathcal{I}})$.
2. $\text{SA}_+^r(R_{\mathcal{I}}) \subseteq \text{SA}_+^{r+k+1}(L_{\mathcal{I}})$ and $\text{SA}_+^r(L_{\mathcal{I}}) \subseteq \text{SA}_+^{r+k+1}(R_{\mathcal{I}})$.
3. $\text{StaticLS}_+^r(R_{\mathcal{I}}) \subseteq \text{StaticLS}_+^{r+k+1}(L_{\mathcal{I}})$ and $\text{StaticLS}_+^r(L_{\mathcal{I}}) \subseteq \text{StaticLS}_+^{r+k+1}(R_{\mathcal{I}})$.

We include the proof in Appendix E. We define the SA^r , SA_+^r , and StaticLS_+^r operators in the next sections.

Larger alphabets We can also consider CSPs with variables taking values in $[q]$. We use $\{0, 1\}$ -valued variables $x_{i,a}$ such that $x_{i,a} = 1$ if and only if variable i is assigned value a . Given predicate $P : [q]^k \rightarrow \{0, 1\}$, let $P_{01} : \{0, 1\}^{qk} \rightarrow \{0, 1\}$ be a polynomial in variables $z_{i,a}$ such that

$$P_{01}(z) = \sum_{\alpha \in \mathbb{Z}_q^k} P(\alpha) \prod_{i \in [k]} z_{i, \alpha_i}.$$

Observe that $P(\alpha) = P_{01}(z)$ if $z_{i, \alpha_i} = 1$ and $z_{i,b} = 0$ for all $i \in [k]$ and $b \neq \alpha_i$. Given $z \in \mathbb{R}^{qk}$ and $c \in [q]^k$, define $z^{(c)} \in \mathbb{R}^{qk}$ so that $z_{i,a}^{(c)} = z_{i, (a+c_i) \bmod q}$. The constraints in this formulation have degree k .

We can encode the decision problem as the following system of polynomial inequalities.

$$R_{\mathcal{I}} = \left\{ P_{01}(x_{S \times [q]}^{(c)}) = 1 \mid (c, S) \in \mathcal{I} \right\}.$$

2.2.2 Sherali-Adams

The Sherali-Adams (SA) linear programming hierarchy gives a family of locally consistent distributions on assignments to sets of variables. As the size of these sets increases, the relaxation becomes tighter.

Definition 2.4. *Let $\{D_S\}$ be a family of distributions D_S over $[q]^S$ for all $S \subseteq [n]$ with $|S| \leq r$. We say that $\{D_S\}$ is r -locally consistent if for all $T \subseteq S \subseteq [n]$ with $|S| \leq r$, the marginal of D_S on T is equal to D_T . In symbols, we write this condition as $D_T(T = \alpha) = D_S(T = \alpha)$, where*

$$D_S(T = \alpha) = \sum_{\substack{\beta \in [q]^S \\ \beta_T = \alpha}} D_S(\beta).$$

We can now define the SA relaxation of a set of constraints. We will extend the distributions $\{D_S\}$ to distributions over assignments to x in $[q]^n$ by choosing assignments to $x_{[n]\setminus S}$ uniformly at random. Given a family of r -locally consistent distributions $\{D_S\}$ and a monomial $x^T = \prod_{i \in T} x_i$ with $|T| \leq r$, we define $\mathbb{E}_D[x^T] = \mathbb{E}_{D_T}[x^T]$. We extend this definition to degree- r polynomials by linearity. For $T \subseteq [n]$ and $\alpha \in [q]^{|T|}$, let $1_{\{x_T=\alpha\}}(x)$ be the indicator polynomial for the event $x_T = \alpha$. Let $\deg(\cdot)$ denote the multilinear degree of a polynomial, i.e., the degree after replacing all appearances of x_i^2 with x_i .

Definition 2.5. Let $A = \{g_1(x) \geq 0, g_2(x) \geq 0, \dots, g_m(x) \geq 0\}$ be a set of polynomial constraints such that for all $g \in A$, $\deg(g) \leq r$. We define the r -round SA relaxation for A to be the set of all families of distributions $\{D_S\}_{S \subseteq [n], |S| \leq r}$ such that D_S is a distribution over $[q]^S$ satisfying the following two properties.

1. $\{D_S\}_{S \subseteq [n], |S| \leq r}$ is r -locally consistent.
2. $\mathbb{E}_{x \sim D}[1_{\{x_T=\alpha\}}(x) \cdot g(x)] \geq 0$ for all $g \in A$, $T \subseteq [n]$, and $\alpha \in [q]^{|T|}$ such that $\deg(1_{\{x_T=\alpha\}}(x) \cdot f(x)) \leq r$.

We denote this set of families of distributions as $\text{SA}^r(A)$. $\text{SA}^r(A)$ is a polytope of size $n^{O(r)}$ and we can therefore check feasibility in time $n^{O(r)}$. We point out that $\text{SA}^{r+1}(A) \subseteq \text{SA}^r(A)$ and $\text{SA}^n(A)$ exactly captures feasibility of A over $[q]^n$.

Specialized to an instance \mathcal{I} of $\text{CSP}(P)$, we consider $\text{SA}^r(R_{\mathcal{I}})$ and write Condition 2 as

$$\mathbb{E}_{x \sim D}[1_{\{x_T=\alpha\}}(x) \cdot (P(x_S + c) - 1)] = 0$$

for all $(c, S) \in \mathcal{I}$, all $T \subseteq [n]$ such that $\deg(1_{\{x_T=\alpha\}}(x) \cdot (P(x_S + c) - 1)) \leq r$, and all $\alpha \in [q]^{|T|}$.

We will only consider the stronger feasibility formulations of relaxations (rather than optimization versions) because lower bounds for these feasibility formulations immediately imply lower bounds for the corresponding optimization versions.

For larger alphabets, implementing this definition as a linear program requires writing the constraints as polynomials in Boolean variables. We therefore identify $\text{SA}^r(A)$ with $\text{SA}^r(A_{01})$, where the A_{01} is the inequalities of A written as polynomials in the Boolean variables $x_{i,a}$ as described above. We make this same identification for SA_+ and static LS_+ below.

In previous work, O'Donnell and Witmer [OW14] extended a theorem of Benabbas et al. [BGMT12] to obtain a lower bound for SA relaxations of $\text{CSP}(P)$ with $m = n^{t/2-\varepsilon}$ when P is $(t-1)$ -wise uniform supporting.

Theorem 2.6 ([BGMT12, OW14]). Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t-1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability $\text{SA}^r(R_{\mathcal{I}})$ is nonempty for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.

As mentioned above, feasibility of the SA relaxation of the decision version of $\text{CSP}(P)$ immediately implies that the optimization version of the SA relaxation has value 1, i.e., SA thinks all constraints can be satisfied. The same holds for the other relaxations we consider; we only look at feasibility for the rest of the paper.

2.2.3 Sherali-Adams₊

The Sherali-Adams₊ (SA_+) SDP hierarchy additionally requires the second moment matrix of these distributions to be PSD. Given a family of local distributions $\{D_S\}$, define $M = M(D) \in \mathbb{R}^{(nq+1) \times (nq+1)}$ to be the symmetric matrix indexed by $(0, [n] \times [q])$ such that

$$\begin{aligned} M(0, 0) &= 1 \\ M(0, (i, a)) &= D_{\{i\}}(x_i = a) \\ M((i, a), (j, b)) &= D_{\{i, j\}}(x_i = a \wedge x_j = b). \end{aligned}$$

Note that the $((i, a), (i, b))$ -element of M is $D_{\{i\}}(x_i = a)$ if $a = b$ and is 0 if $a \neq b$.

Definition 2.7. Given a set of constraints A as above, we define $\text{SA}_+^r(A)$ to be the set of all families of distributions $\{D_S\}_{S \subseteq [n], |S| \leq r}$ over $[q]^S$ in $\text{SA}^r(A)$ satisfying the following additional condition.

3. M is PSD.

We can equivalently define SA_+ by requiring the covariance matrix of the locally consistent $\{D_S\}$ distributions to be positive semidefinite.

Definition 2.8. Given r -locally consistent distributions $\{D_S\}$ with $r \geq 2$, the covariance matrix $\Sigma = \Sigma(D)$ is defined to be

$$\Sigma((i, a), (j, b)) = D_{\{i, j\}}(x_i = a \wedge x_j = b) - D_{\{i\}}(x_i = a) \cdot D_{\{j\}}(x_j = b).$$

These two formulations are equivalent [WJ08].

Lemma 2.9. M is PSD if and only if Σ is PSD.

We include the proof in Appendix B. The covariance matrix condition will be more convenient for us to work with. For an instance \mathcal{I} of $\text{CSP}(P)$, we will consider feasibility of $\text{SA}_+^r(R_{\mathcal{I}})$. We prove the following theorem.

Theorem 1.3 (restated). Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t-1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability $\text{SA}_+^r(R_{\mathcal{I}})$ is nonempty for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.

2.2.4 Lovász-Schrijver $_+$

We now define the Lovász-Schrijver $_+$ (LS_+) SDP relaxation for problems whose variables are 0/1-valued. Given an initial polytope $K \in \mathbb{R}^n$, we would like to generate a sequence of progressively tighter relaxations. To define one LS_+ lift-and-project step, we will use the cone

$$\tilde{K} = \{(\lambda, \lambda x_1, \dots, \lambda x_n) \mid \lambda > 0, x_1, \dots, x_n \in K\}.$$

K can be recovered by taking the intersection with $x_0 = 1$.

Definition 2.10. Let \tilde{K} be a convex cone in \mathbb{R}^{n+1} . Then the lifted LS_+ cone $N_+(\tilde{K})$ is the cone of all $y \in \mathbb{R}^{n+1}$ for which there exists an $(n+1) \times (n+1)$ matrix Y satisfying the following:

1. Y is symmetric and positive semidefinite.
2. For all i , $Y_{ii} = Y_{i0} = y_i$.
3. For all i , $Y_i \in \tilde{K}$ and $Y_0 - Y_i \in \tilde{K}$

where Y_i is the i th column of Y . Then we define $N_+(K)$ to be $N_+(\tilde{K}) \cap \{x_0 = 1\}$. The r -round LS_+ relaxation of a polytope K results from applying the N_+ operator r times. That is, we define $N_+^r(K) = N_+(N_+^{r-1}(K))$. Y is called a protection matrix for y . A solution to the r -round LS_+ relaxation for a polytope $K \in \mathbb{R}^n$ defined by $\text{poly}(n)$ linear constraints can be computed in time $n^{O(r)}$ using an SDP.

For an instance \mathcal{I} of $\text{CSP}(P)$, we will consider feasibility of $N_+^r(L_{\mathcal{I}})$.

Theorem 1.4 (restated). Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t-1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability $N_+^r(L_{\mathcal{I}})$ is nonempty for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.

2.2.5 Static LS_+

The static LS_+ relaxation strengthens both SA_+ and LS_+ . As in the case of SA_+ , we start with a family of r -locally consistent distributions and then further require that they satisfy certain positive semidefiniteness constraints. In particular, for all $X \subseteq [n]$ with $|X| \leq r - 2$ and all $\alpha \in [q]^X$, define the matrices $M_{X,\alpha} = M_{X,\alpha}(D) \in \mathbb{R}^{(nq+1) \times (nq+1)}$ as follows.

$$\begin{aligned} M_{X,\alpha}(0, 0) &= D_X(X = \alpha) \\ M_{X,\alpha}(0, (i, a)) &= D_{\{i\} \cup X}(x_i = a \wedge X = \alpha) \\ M_{X,\alpha}((i, a), (j, b)) &= D_{\{i,j\} \cup X}(x_i = a \wedge x_j = b \wedge X = \alpha). \end{aligned}$$

In addition to the SA constraints, the r -round static LS_+ relaxation $\text{StaticLS}_+^r(F)$ satisfies the following constraint.

Definition 2.11. *Given a set of constraints A as above, we define $\text{StaticLS}_+^r(A)$ to be the set of all families of distributions $\{D_S\}_{S \subseteq [n], |S| \leq r}$ over $[q]^S$ in $\text{SA}^r(A)$ satisfying the following additional condition.*

3'. $M_{X,\alpha}$ is PSD for all $X \subseteq [n]$ with $|X| \leq r - 2$ and all $\alpha \in [q]^X$.

Observe that these positive semidefiniteness constraints can be formulated as a positive semidefiniteness constraint for a single matrix. In particular, let \mathcal{M} be the block diagonal matrix with the $M_{X,\alpha}$'s on the diagonal. Then \mathcal{M} has size at most $(qn)^{O(r)}$ and is PSD if and only if all of the $M_{X,\alpha}$'s are PSD. Unlike LS_+ , this hierarchy easily generalizes to non-binary alphabets.

Alternatively, we can think of this hierarchy as requiring covariance matrices of conditional distributions to be positive semidefinite. Given a set of local distributions $\{D_S\}$, a set of variables $X \subseteq [n]$, and an assignment $\alpha \in \{0, 1\}^X$ such that $D_X(X = \alpha) > 0$, define a set of conditional local distributions $\{D_{S|X=\alpha}\}$ by $D_{S|X=\alpha}(\beta) = \frac{D_{S \cup X}(S=\beta \wedge X=\alpha)}{D_X(X=\alpha)}$. Tulsiani and Worah showed that if the initial family of local distributions is r -locally consistent, then the corresponding family of conditional distributions will be $(r - |X|)$ -locally consistent [TW13].

Lemma 2.12 ([TW13, Lemma 3.13]). *Let $X \subseteq [n]$ and let $\{D_S\}$ be a family of r -locally consistent distributions for sets $S \subseteq [n]$ such that $S \cap X = \emptyset$ and $|S \cup X| \leq r$. Then the family of conditional distributions $\{D_{S|X=\alpha}\}$ is $(r - |X|)$ -locally consistent for any $\alpha \in \{0, 1\}^X$ such that $D_X(X = \alpha) > 0$.*

We include the proof of this lemma in Appendix C.

Given such a family of conditional local consistent distributions, the conditional covariance matrix $\Sigma_{X,\alpha}$ is defined as follows.

Definition 2.13. *Given $X \subseteq [n]$, $\alpha \in \{0, 1\}^X$, and r -locally consistent conditional distributions $\{D_{S|X=\alpha}\}$ with $r \geq 2$, the conditional covariance matrix $\Sigma_{X,\alpha} = \Sigma_{X,\alpha}(D)$ is defined to be*

$$\Sigma_{X,\alpha}((i, a), (j, b)) = \begin{cases} D_{\{i,j\}|X=\alpha}(x_i = a \wedge x_j = b) - D_{\{i\}|X=\alpha}(x_i = a) \cdot D_{\{j\}|X=\alpha}(x_j = b) & \text{if } D_X(\alpha) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2.9 generalizes to these conditional covariance matrices.

Lemma 2.14. $M_{X,\alpha}$ is PSD if and only if $\Sigma_{X,\alpha}$ is PSD.

The proof is essentially identical to that of Lemma 2.9.

We note that we have not seen this hierarchy defined in this form in previous work, but it is dual to the static LS_+ proof system defined in [GHP02] and described below in Section 2.3 (see Proposition F.1). For a random instance \mathcal{I} of $\text{CSP}(P)$, we will study feasibility of $\text{StaticLS}_+^r(R_{\mathcal{I}})$.

Theorem 1.5 (restated). Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability $\text{StaticLS}_+^r(R_{\mathcal{I}})$ is nonempty for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ rounds.

2.3 The dual point of view: Semialgebraic proof systems

We can also consider refutation of CSPs via semialgebraic proof systems. Starting from a set of polynomial inequalities $A = \{g_1(x) \geq 0, g_2(x) \geq 0, \dots, g_m(x) \geq 0\}$ called axioms that encode the constraints of the CSP, semialgebraic proof systems derive new inequalities that are implied by A . To prove that an instance is unsatisfiable, we wish to derive the contradiction $-1 \geq 0$. We consider the SA, SA_+ , LS_+ , and static LS_+ proof systems. In this section, we again think of $1_{\{x_T=\alpha\}}$ as a polynomial. We give definitions for binary alphabets. For larger alphabets, we can rewrite constraints in terms of binary variables as described above. When refuting a CSP \mathcal{I} , we start with constraints $A_{\mathcal{I}}$ for SA, SA_+ , and static LS_+ and use constraints $L_{\mathcal{I}}$ in the case of LS_+ .

The SA proof system An SA refutation of A has the form

$$\sum_{g \in A} g(x) \sum_i \gamma_{g,i} \cdot 1_{\{x_{T_{g,i}}=\alpha_{g,i}\}}(x) + \sum_j (x_j^2 - x_j) h_j(x) = -1,$$

where $\gamma_{g,i} \geq 0$ and the h_j 's are arbitrary polynomials. This is a proof of unsatisfiability because under the assumption that the all x_i variables are in $\{0, 1\}$, every term of the above sum must be nonnegative and it is therefore a contradiction. The *degree* of this proof is the maximum degree of any of the terms. The *size* of the proof is the number of terms in the sum. An degree- r SA refutation exists if and only if the corresponding r -round SA relaxation is unsatisfiable; this follows from Farkas' Lemma. The SA proof system is automatizable: A degree- r SA refutation may be found in time $n^{O(r)}$ by solving an LP if it exists. The SA proof system first appeared in [GHP02] with the name static LS^∞ ; the dual hierarchy of LP relaxations was introduced by [SA90].

Lower bounds of Benabbas et al. [BGMT12] and O'Donnell and Witmer [OW14] immediately imply that there are no degree- $n^{\varepsilon/(t-2)}$ SA refutations for random instances of $\text{CSP}(P)$ with $(t - 1)$ -wise uniform supporting P and $m \leq n^{t/2-\varepsilon}$.

Corollary 2.15. Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability there is no degree- r SA refutation of $A_{\mathcal{I}}$ for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.

The SA_+ proof system In SA_+ , a proof has the form

$$\sum_{f \in A} g(x) \sum_i \gamma_{g,i} \cdot 1_{\{x_{T_{g,i}}=\alpha_{g,i}\}}(x) + \sum_j (x_j^2 - x_j) h_j(x) + \sum_\ell \eta_\ell^2(x) = -1$$

where $\gamma_{g,i} \geq 0$, the h_j 's are arbitrary polynomials, and the η_ℓ 's are affine functions. The dual SA_+ hierarchy of SDP relaxations first appeared in [Rag08]. Again, a degree- r SA_+ refutation exists if and only if the corresponding r -round SA_+ relaxation is infeasible. We do not know any of any results on the automatizability of SA_+ .

Our lower bound for SA_+ SDP relaxations of random instances of $\text{CSP}(P)$ implies a lower bound on the degree of SA_+ refutations of random instances of $\text{CSP}(P)$.

Corollary 2.16. Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $\text{CSP}(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability there is no degree- r SA_+ refutation of $A_{\mathcal{I}}$ for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.

The LS_+ proof system The LS_+ proof system [LS91] is dynamic, meaning that a proof is built up over a series of steps. A proof in LS_+ is a sequence of polynomial inequalities $Q(x) \geq 0$. A new inequality is derived from the inequalities already in the proof using inference rules. When $\deg(Q(x)) \leq 1$, we allow the following:

$$\frac{Q(x) \geq 0}{x_i \cdot Q(x) \geq 0} \quad \frac{Q(x) \geq 0}{(1 - x_i) \cdot Q(x) \geq 0} \quad \frac{}{Q(x)^2 \geq 0}.$$

We also allow nonnegative linear combinations:

$$\frac{Q(x) \geq 0 \quad R(x) \geq 0}{\alpha \cdot Q(x) + \beta \cdot R(x) \geq 0}$$

for $\alpha, \beta \geq 0$. An LS_+ proof is therefore a sequence of “lifting” steps in which we multiply by some x_i or $(1 - x_i)$ to get a degree-2 polynomial and “projection” steps in which we take nonnegative linear combinations to reduce the degree back to 1. We can view an LS_+ proof as a directed acyclic graph with inequalities at each vertex and $-1 \geq 0$ at the root. The *rank* of an LS_+ proof is the maximum number of lifting steps in any path to the root. A rank- r LS_+ refutation exists if and only if the corresponding rank- r LS_+ relaxation is infeasible [Das01]. The LS_+ proof system is not known to be automatizable; see Section 8 of [BOGH⁺06] for details.

Our LS_+ lower bound implies a lower bound on the rank of LS_+ refutations of random instances of $CSP(P)$.

Corollary 2.17. *Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $CSP(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability there is no rank- r LS_+ refutation of $L_{\mathcal{I}}$ for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.*

The static LS_+ proof system A static LS_+ proof [GHP02] has the following form.

$$\sum_i \gamma_i \cdot b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x) + \sum_j (x_j^2 - x_j) h_j(x) = -1,$$

where $\gamma_i \geq 0$, either $b_i \in A$ or $b_i = \eta_i^2$ for some affine function η_i , and the h_j ’s are arbitrary polynomials. Note that this proof system is at least as powerful as the SA_+ proof system: Terms in the sum may be products of a $1_{\{x_T=\alpha\}}$ term and the square of an affine function instead of just the square of an affine function or just an axiom multiplied by a $1_{\{x_T=\alpha\}}$ term. Once again, there exists a static LS_+ refutation if and only if the corresponding static LS_+ relaxation is infeasible. We do not know of any proof of this statement in the literature, so we include one in Appendix F. We do not know of any results on the automatizability of static LS_+ .

Again, our static LS_+ lower bound implies a lower bound on the degree of static LS_+ refutations of $CSP(P)$.

Corollary 2.18. *Let $P : [q]^k \rightarrow \{0, 1\}$ be $(t - 1)$ -wise uniform-supporting and let \mathcal{I} be a random instance of $CSP(P)$ with n variables and $m \leq n^{t/2-\varepsilon}$ constraints. Then with high probability there is no degree- r LS_+ refutation of $A_{\mathcal{I}}$ for $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.*

2.4 Expansion

Given a set of constraints T , we define its neighbor set $\Gamma(T) \subseteq [n]$ as $\Gamma(T) = \{v \in [n] \mid v \in S \text{ for some } (c, S) \in T\}$. We can then define expansion.

Definition 2.19. *An instance \mathcal{I} of $CSP(P)$ is (s, e) -expanding if for every set of constraints T with $|T| \leq s$, $|\Gamma(T)| \geq e|T|$.*

We can also define T ’s boundary neighbors as $\partial T = \{v \in [n] \mid v \in S \text{ for exactly one } (c, S) \in T\}$ and define a corresponding notion of boundary expansion.

Definition 2.20. An instance \mathcal{I} of $\text{CSP}(P)$ is (s, e) -boundary expanding if for every set of constraints T with $|T| \leq s$, $|\partial T| \geq e|T|$.

We state a well-known connection between expansion and boundary expansion appearing in, e.g., [TW13].

Fact 2.21. $(s, k - d)$ -expansion implies $(s, k - 2d)$ -boundary expansion.

It is also well-known that randomly-chosen sets of constraints have high expansion [BGMT12, OW14]:

Lemma 2.22. Fix $\varepsilon > 0$. With high probability, a set of $m = n^{t/2-\varepsilon}$ constraints chosen uniformly at random is both $(n^{\frac{\varepsilon}{t-2}}, k - \frac{t}{2} + \frac{\varepsilon}{2})$ -expanding and $(n^{\frac{\varepsilon}{t-2}}, k - t + \varepsilon)$ -boundary expanding.

We give proofs of both of these statements in Appendix A.

2.5 Constructing consistent local distributions

Here, we recall a construction of consistent local distributions supported on satisfying assignments. We will study these distributions in the remainder of this paper, showing that they are valid SA_+ , LS_+ , and static LS_+ solutions. They were first used in [BGMT12] and have appeared in many subsequent works (e.g., [TW13, OW14, BCK15]). In Appendix D, we give proofs of all results mentioned in this section.

We first need to mention the notion of a closure of a set of variables. For $S \subseteq [n]$, let $H_{\mathcal{I}} - S$ denote the hypergraph $H_{\mathcal{I}}$ with the vertices of S and all hyperedges contained in S removed. Intuitively, the closure $\text{Cl}(S)$ of a set $S \subseteq [n]$ is a superset of S that is not too much larger than S and is not very well-connected to the rest of the instance in the sense that $H_{\mathcal{I}} - S$ has high expansion.

Lemma 2.23 ([BGMT12, TW13]). If $H_{\mathcal{I}}$ is (s_1, e_1) -expanding and S is a set of variables such that $|S| < (e_1 - e_2)s_1$ for some $e_2 \in (0, e_1)$, then there exists a set $\text{Cl}(S) \subseteq [n]$ such that $S \subseteq \text{Cl}(S)$ and $H_{\mathcal{I}} - \text{Cl}(S)$ is (s_2, e_2) -expanding with $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$ and $|\text{Cl}(S)| \leq \frac{k + 2e_1 - e_2}{2(e_1 - e_2)} |S|$.

We give a formal definition of the closure and a proof of this lemma in Appendix D.

We now use the closure to define consistent local distributions supporting on satisfying assignments. We assume that there exists a $(t - 1)$ -wise independent distribution μ over satisfying assignments to P . For a constraint $C = (c, S)$, let μ_C be the distribution defined by $\mu_C(z) = \mu(z_1 + c_1, \dots, z_k + c_k)$ and let $\mathcal{C}(S)$ be the set of constraints whose support is entirely contained within S . For a set of variables $S \subseteq [n]$ and an assignment $\alpha \in [q]^S$, we use the notation $S = \alpha$ to indicate the the variables of S are labeled according to the assignment α . For a constraint $C = (c, S)$ and an assignment α to a superset of S , let $\mu_C(\alpha) = \mu_C(\alpha_S)$.

For $S \subseteq [n]$, we can then define the distribution Π'_S over $[q]^S$ as

$$\Pi'_S(S = \alpha) = \frac{1}{Z_S} \prod_{C \in \mathcal{C}(S)} \mu_C(\alpha), \text{ where } Z_S = \sum_{\beta \in [q]^S} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta).$$

Using $\{\Pi'_S\}$, we can then define local distributions Π_S by $\Pi_S(S = \alpha) = \Pi'_{\text{Cl}(S)}(S = \alpha)$. [BGMT12, OW14] proved that these distributions are r -locally consistent for $r = n^{\frac{\varepsilon}{t-2}}$.

Theorem 2.24 ([BGMT12, OW14]). For a random instance \mathcal{I} with $m \leq n^{t/2-\varepsilon}$, the family of distributions $\{\Pi_S\}_{|S| \leq r}$ is r -locally consistent for $r = \Omega\left(n^{\frac{\varepsilon}{t-2}}\right)$ and is supported on satisfying assignments.

This theorem shows that the SA cannot efficiently refute random $(t - 1)$ -wise uniforming supporting instances: the r -round SA LP still has value 1 for some $r = \Omega(n^{\frac{\varepsilon}{t-2}})$ when $m \leq n^{t/2-\varepsilon}$. In this paper, we show that even when we add the SA_+ requirement that the covariance matrix is PSD, we still cannot refute when $m \leq n^{t/2-\varepsilon}$.

Given these locally consistent distributions, Lemma 2.12 implies that the conditional distributions $\{\Pi_{S|X=\alpha}\}$ defined above are also locally consistent.

Corollary 2.25 ([TW13]). *There exists a constant $c > 0$ such that the following holds. Let \mathcal{I} be a random instance of $\text{CSP}(P)$ with $m \leq n^{t/2-\varepsilon}$. Let $X \subseteq [n]$ such that $|X| \leq cn^{\frac{\varepsilon}{t-2}}$ and let $\alpha \in \{0, 1\}^X$ be any assignment to X such that $\mu_C(\alpha) > 0$ for all constraints in $C \in \mathcal{C}(X)$. Then the family of conditional distributions $\{\Pi_S|_{X=\alpha}\}_{|S| \leq r, S \cap X = \emptyset}$ is r -locally consistent for some $r = \Omega(n^{\frac{\varepsilon}{t-2}})$.*

We will use these conditional consistent local distributions to prove lower bounds for LS_+ in Section 7.

3 Overview of the proof

Showing that a set of local distributions is a valid SA_+ solution requires proving that these distributions are locally consistent and proving that their covariance matrix is PSD. Local consistency of the $\{\Pi_S\}$ distributions was proven in previous work [BGMT12, OW14]. To prove Theorem 1.3, it remains to argue that the covariance matrix of $\{\Pi_S\}$ is PSD.

Previous work [BGMT12, TW13] only considers instances with a linear number of constraints and relies on the fact that most pairs of variables are uncorrelated in this regime. For $m \gg n$, however, correlations between pairs of vertices do arise because the underlying hypergraph becomes more dense. The major technical contribution of this work is to deal with these correlations by proving that they remain local. We consider the graph induced by correlations between variables: Two variables are connected if they have non-zero correlation. We prove that this graph must have connected components of at most constant size with high probability. Each of these connected components can then be covered by a local distribution of constant size. This implies that each submatrix of the covariance matrix corresponding to one of these connected component is PSD, and thus the entire covariance matrix is PSD.

The proof of Theorem 1.3 has three steps. First, we show in Section 4 that if the correlation graph has small connected components, then the covariance matrix is PSD. Second, we show that any non-zero correlation must have been caused by a relatively dense subset of constraints in Section 5. In Section 6, we show that connected components in the correlation graph must be small or they would induce large dense subsets of constraints that would violate expansion properties.

In Section 7, we show that this same strategy can be used to prove PSDness of conditional covariance matrices and thereby prove Theorems 1.4 and 1.5.

4 The correlation graph

In this section, we define the correlation graph and show that if the correlation graph has small connected components, then the covariance matrix is PSD.

Definition 4.1. *The correlation graph G_{corr} associated with r -locally consistent distributions $\{D_S\}$ is the graph on $[n]$ with an edge between every pair of variables for which there is a nonzero entry in the covariance matrix for $\{D_S\}$. More formally, the set of edges of G_{corr} is defined to be*

$$E(G_{\text{corr}}) = \{(u, v) \in [n] \times [n] \mid u \neq v, \exists (a, b) \in [q] \times [q] \text{ s.t. } \Sigma_{(u,a),(v,b)} \neq 0\}.$$

Lemma 4.2. *Let $\{D_S\}$ be a family of r -locally consistent distributions. If all connected components in the correlation graph associated with $\{D_S\}$ have size at most r , then the covariance matrix for $\{D_S\}$ is PSD.*

Proof. Consider the partition V_1, V_2, \dots, V_ℓ of $[n]$ such that u and v are in the same set if and only if they are connected in the correlation graph. We then have nonzero entries in the covariance matrix only for pairs $((u, a), (v, b))$ such that $u, v \in V_i$ for some i . Ordering the rows and columns of the covariance matrix according to this partition, we see that the covariance matrix is block diagonal with a nonzero block for each connected component of the correlation graph. Each of these blocks is PSD since each is the covariance matrix of the local distribution D_{V_i} for the corresponding set V_i with size at most r , and the covariance matrix of valid distribution is always PSD. Since each block is PSD, the entire matrix is PSD. \square

We already know that $\{\Pi_S\}$ defined in Section 2.5 is $\Omega(n^{\frac{\epsilon}{t-2}})$ -locally consistent with high probability when $m \leq n^{t/2-\epsilon}$. In the following sections, we will show that connected components in the correlation graph associated with $\{\Pi_S\}$ are small. Hence, from Lemma 4.2, $\{\Pi_S\}$ is a feasible solution for the SA_+ SDP.

5 Correlations are induced by small, dense structures

In this section, we show that pairwise correlations in $\{\Pi_S\}$ are only generated by small, dense subhypergraphs that we will call “bad structures”. Given a set of hyperedges W , call a variable v an W -boundary variable if it is contained in exactly one constraint in W .

Definition 5.1. *For variables u and v , a bad structure for u and v is a set of constraints W satisfying the following properties:*

1. $u, v \in \Gamma(W)$.
2. The hypergraph induced by W is connected.
3. Every constraint contains at most $k - t$ W -boundary variables other than u and v .

We also say W is a bad structure if W is a bad structure for some u and v .

A bad structure for u and v generates correlation between u and v with respect to $\{\Pi_S\}$.

Lemma 5.2. *If there is no bad structure for u and v of size at most $|\mathcal{C}(\text{Cl}(\{u, v\}))|$, then u and v are not correlated with respect to $\Pi_{\{u, v\}}$.*

We need the following technical claim, which states that the distribution Π'_S isn't affected by removing a constraint with many boundary variables.

Claim 5.3. *Let $T \subseteq S \subseteq [n]$ be sets of variables. Let $C^* \in \mathcal{C}(S)$ be some constraint covered by S . If $|(\partial\mathcal{C}(S) \cap C^*) \setminus T| \geq k - t + 1$, then for any $\alpha \in \{0, 1\}^T$,*

$$\Pi'_S(T = \alpha) \propto \Pi'_{S \setminus (\partial\mathcal{C}(S) \cap C^*)}(T \setminus (\partial\mathcal{C}(S) \cap C^*) = \alpha_{T \setminus (\partial\mathcal{C}(S) \cap C^*)}),$$

Proof. Let $B = \partial\mathcal{C}(S) \cap C^*$ be the boundary variables of $\mathcal{C}(S)$ contributed by C^* , i.e., the variables contained in C^* that don't appear in any other constraint of $\mathcal{C}(S)$. Then

$$\begin{aligned} \Pi'_S(T = \alpha) &\propto \sum_{\substack{\beta \in \{0, 1\}^S \\ \beta_T = \alpha}} \prod_{C \in \mathcal{C}(S)} \mu_C(\beta) \\ &= \sum_{\substack{\beta \in \{0, 1\}^{S \setminus B} \\ \beta_{T \setminus B} = \alpha_{T \setminus B}}} \prod_{C \in \mathcal{C}(S) \setminus \{C^*\}} \mu_C(\beta) \sum_{\substack{\gamma \in \{0, 1\}^B \\ \gamma_{B \cap T} = \alpha_{B \cap T}}} \mu_{C^*}(\beta, \gamma) \\ &= \frac{1}{q^{k - |B \setminus T|}} \sum_{\substack{\beta \in \{0, 1\}^{S \setminus B} \\ \beta_{T \setminus B} = \alpha_{T \setminus B}}} \prod_{C \in \mathcal{C}(S) \setminus \{C^*\}} \mu_C(\beta) \\ &\propto \Pi'_{S \setminus B}(T \setminus B = \alpha_{T \setminus B}). \end{aligned}$$

The second-to-last line holds because $|B \setminus T| \geq k - t + 1$ and μ is $(t - 1)$ -wise independent. □

Using Claim 5.3, we prove Lemma 5.2.

Proof of Lemma 5.2. Let $S_0 = \mathcal{C}(\text{Cl}(\{u, v\}))$. Say there exists a constraint C_1 such that $|(\partial\mathcal{C}(S_0) \cap C_1) \setminus \{u, v\}| \geq k-t+1$. Let $S_1 = S_0 \setminus (\partial\mathcal{C}(S_0) \cap C_1)$. If there exists a constraint C_2 such that $|(\partial\mathcal{C}(S_1) \cap C_2) \setminus \{u, v\}| \geq k-t+1$, remove its boundary variables in the same manner to get S_2 . Continue in this way until we obtain a set S_ℓ such that $|(\partial\mathcal{C}(S_\ell) \cap C) \setminus \{u, v\}| \leq k-t$ for every constraint $C \in \mathcal{C}(S_\ell)$ ($\mathcal{C}(S_\ell)$ could be empty). Since $|(\partial\mathcal{C}(S_{i-1}) \cap C_i) \setminus \{u, v\}| \geq k-t+1$ for $1 \leq i \leq \ell$, we can apply Claim 5.3 ℓ times to see that

$$\Pi_{\{u,v\}}(u=a \wedge v=b) \propto \begin{cases} \Pi'_{S_\ell}(u=a) & \text{if } u \in S_\ell, v \notin S_\ell \\ \Pi'_{S_\ell}(v=b) & \text{if } v \in S_\ell, u \notin S_\ell \\ 1 & \text{if } u, v \notin S_\ell \\ \Pi'_{S_\ell}(u=a \wedge v=b) & \text{if } u, v \in S_\ell. \end{cases}$$

In the first three cases, it is easy to see that the lemma holds. In the last case, we know that $\mathcal{C}(S_\ell)$ cannot be a bad structure by our assumption. Since $\mathcal{C}(S_\ell)$ satisfies Conditions 1 and 3 of Definition 5.1, the hypergraph induced by S_ℓ must be disconnected with u and v in different connected components. Say S_u and S_v are the vertex sets of the connected components of S_ℓ containing u and v , respectively. Then

$$\Pi_{\{u,v\}}(u=a \wedge v=b) \propto \Pi'_{S_\ell}(u=a \wedge v=b) = \Pi'_{S_u}(u=a) \cdot \Pi'_{S_v}(v=b).$$

The result then follows. \square

6 All connected components of the correlation graph are small

In this section, we show that all connected components in the correlation graph associated with $\{\Pi_S\}$ are small, which concludes the proof of Theorem 1.3.

Theorem 6.1. *Assume that the hypergraph $H_{\mathcal{I}}$ is an $(r, k-t/2 + \delta/2)$ -expander for some $r = \omega(1)$. Then all connected components in the correlation graph associated with $\{\Pi_S\}$ have size at most $\frac{2k}{\delta}$.*

We will actually prove a slightly more general theorem that we will use to prove LS_+ lower bounds in Section 7. Given a hypergraph H , let $G_{\text{bad}}(H)$ be the graph on $[n]$ such that there is an edge between i and j if and only if there exists a bad structure for i and j in H .

Theorem 6.2. *If the hypergraph H is an $(r, k-t/2 + \delta/2)$ -expander for some $r = \omega(1)$, then all connected components in $G_{\text{bad}}(H)$ have size at most $\frac{2k}{\delta}$.*

Lemma 5.2 implies that $G_{\text{bad}}(H_{\mathcal{I}})$ contains the correlation graph associated with $\{\Pi_S\}$ as a subgraph, so Theorem 6.2 immediately implies Theorem 6.1.

Proof of Theorem 6.2. For any edge e of $G_{\text{bad}}(H)$, we can find a corresponding bad structure W_e . We will say that W_e induces e . Any such bad structure W_e satisfies

$$\Gamma(W_e) \leq (k-t)|W_e| + 2 + \frac{k|W_e| - ((k-t)|W_e| + 2)}{2} = \left(k - \frac{t}{2}\right)|W_e| + 1. \quad (6.4)$$

The first term upper bounds the number of boundary vertices, the second term counts the endpoints of e , and the last term upper bounds the number of non-boundary vertices. For a connected component in $G_{\text{bad}}(H)$, let e_1, e_2, \dots, e_ℓ be an ordering of edges in the connected component such that $(\bigcup_{j=1}^i e_j) \cap e_{i+1}$ is not empty for $i = 1, \dots, \ell$. That is, e_1, e_2, \dots, e_ℓ is an ordering of the edges in the connected component such that every edge except for the first one is adjacent to some edge preceding it. Let $W_{e_1}, \dots, W_{e_\ell}$ be corresponding bad structures inducing these edges. Let $T_i = \bigcup_{j=1}^i W_{e_j}$ for $i = 1, \dots, \ell$. While T_i itself is not necessarily a bad structure, we will show that the inequality (6.4) still holds for T_i , i.e.,

$$\Gamma(T_i) \leq \left(k - \frac{t}{2}\right)|T_i| + 1 \quad (6.5)$$

for any $i = 1, \dots, \ell$. If (6.5) holds, the number of constraints in T_ℓ is at most $\frac{2}{\delta}$; otherwise, expansion is violated. Hence, at most $\frac{2k}{\delta}$ vertices are included in the connected component of the correlation graph associated with $\{D_S\}$.

In the following, we prove (6.5). First, note that $|\Gamma(T_1)| \leq (k - \frac{t}{2})|T_1| + 1$ by (6.4). Let $W'_i = W_{e_i} \setminus T_{i-1}$ be the new constraints added at step i . Call any vertex in $\Gamma(T_i) \setminus \Gamma(T_{i-1})$ a new vertex. We will prove that at most $(k - \frac{t}{2})|W'_i|$ new vertices are added and this will imply (6.5).

Let n_i be the number of new W'_i -boundary vertices. Then the total number of new vertices is at most

$$n_i + (k|W'_i| - 1 - n_i)/2. \quad (6.6)$$

The second term upper bounds the number of non-boundary vertices. The -1 comes from the fact that $\Gamma(W'_i)$ must intersect $\Gamma(T_{i-1})$ since e_i must be adjacent to some preceding edge. If $\Gamma(W'_i)$ and $\Gamma(T_{i-1})$ intersect in a boundary vertex, the resulting bound is stronger.

Hence, we would like to upper bound n_i . We know that n_i is at most $(k - t)|W'_i| + 1$ since any new W'_i -boundary vertex must be a new W_{e_i} -boundary vertex, all but one constraint in W'_i have at most $k - t$ new W_{e_i} -boundary vertices, and one constraint in W'_i has at most $k - t + 1$ new W_{e_i} -boundary vertices. Plugging this into (6.6), we see that the number of new vertices is at most $(k - t)|W'_i| + 1 + (k|W'_i| - 1 - ((k - t)|W'_i| + 1))/2 = (k - t/2)|W'_i|$. \square

From Lemmas 2.22 and 4.2 and Theorems 2.24 and 6.1, we obtain Theorem 1.3.

7 LS_+ rank lower bounds

In this section, we use techniques from the previous sections to prove PSDness of the moment matrices $M_{X,\alpha}$ of the conditional local distributions $\{\Pi_{S|X=\alpha}\}$. From here, degree lower bounds for the static LS_+ proof system and rank lower bounds for LS_+ follow easily.

Lemma 7.1. *There exists a constant $c > 0$ such that the following holds. Let $X \subseteq [n]$ such that $|X| \leq cn^{\frac{\delta}{t-2}}$. For any $\alpha \in [q]^X$ such that $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$, $M_{X,\alpha}$ is positive semidefinite.*

Since we already know that $\{\Pi_S\}$ is a valid SA solution for $|S| = \Omega(n^{\frac{\delta}{t-2}})$, this lemma immediately implies Theorem 1.5. Theorem 1.4, our rank lower bound for LS_+ refutations, follows from Theorem 1.5 and the following fact.

Fact 7.2. *If there exists a rank- r LS_+ refutation of a set of axioms A , then there exists a static LS_+ refutation of A with degree at most r .*

Proof. Let R be a rank- r LS_+ refutation. We look at R as a directed acyclic graph in which each node is the application of some inference rule, the root is $-1 \geq 0$, and the leaves are axioms or applications of the rule $h(x)^2 \geq 0$ for some h with degree at most 1. Starting from the leaves and working back to the root $-1 \geq 0$, we can substitute in the premises of each inference to get an expression $Q(x) = -1$. Since R has rank r , each path in R has at most r multiplications by a term of the form x_i or $(1 - x_i)$ and $Q(x) = -1$ must be a valid static LS_+ refutation of degree at most r . \square

To prove Lemma 7.1, we first show that $M_{X,\alpha}$ is PSD when $H - X$ has high expansion. Then we show that any $M_{X,\alpha}$ can be expressed as a nonnegative combination of $M_{\text{Cl}(X),\beta}$'s for $\beta \in [q]^{\text{Cl}(X)}$. Since $H - \text{Cl}(X)$ has high expansion when $|X| \leq cn^{\frac{\delta}{t-2}}$, each of the $M_{\text{Cl}(X),\beta}$'s is PSD. $M_{X,\alpha}$ is therefore a nonnegative combination of PSD matrices and must itself be PSD.

We start by generalizing Lemma 5.2 to conditional distributions.

Lemma 7.3. *Let $X \subseteq [n]$ and $\alpha \in [q]^X$ such that $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$. If there is no bad structure for u and v in $H - X$ of size at most $|\mathcal{C}(\text{Cl}(\{u, v\}) \setminus X)|$, then u and v are not correlated with respect to $\Pi_{\{u,v\}|X=\alpha}$.*

Proof. First, recall that

$$\Pi_{\{u,v\}|X=\alpha}(u = a \wedge v = b) = \frac{\Pi_{\{u,v\} \cup X}(u = a \wedge v = b \wedge X = \alpha)}{\Pi_X(X = \alpha)}.$$

We will show that $\Pi_{\{u,v\} \cup X}(u = a \wedge v = b \wedge X = \alpha)$ is equal to the product of a term depending on u and a but not v and b and a term depending on v and b but not u and a . From there, the lemma immediately follows.

The proof is essentially the same as that of Lemma 5.2 above. Starting with $S_0 = \mathcal{C}(\text{Cl}(\{u, v\}))$, we apply the same process except we require that each constraint C_i that we remove satisfies $|(\partial\mathcal{C}(S_{i-1}) \cap C_i) \setminus (\{u, v\} \cup X)| \geq k - t + 1$. At the end of this process, we are left with a set S_ℓ such that $|(\partial\mathcal{C}(S_\ell) \cap C) \setminus (\{u, v\} \cup X)| \leq k - t$ for every constraint $C \in \mathcal{C}(S_\ell)$ (again, $\mathcal{C}(S_\ell)$ could be empty). Let $X_\ell = X \cap \Gamma(S_\ell)$ and let $\alpha_\ell = \alpha_{X_\ell}$. By applying Lemma 5.3 repeatedly, we see that

$$\Pi_{\{u,v\} \cup X}(u = a \wedge v = b \wedge X = \alpha) \propto \begin{cases} \Pi'_{S_\ell}(u = a \wedge X_\ell = \alpha_\ell) & \text{if } u \in S_\ell, v \notin S_\ell \\ \Pi'_{S_\ell}(v = b \wedge X_\ell = \alpha_\ell) & \text{if } v \in S_\ell, u \notin S_\ell \\ \Pi'_{S_\ell}(X_\ell = \alpha_\ell) & \text{if } u, v \notin S_\ell \\ \Pi'_{S_\ell}(u = a \wedge v = b \wedge X_\ell = \alpha_\ell) & \text{if } u, v \in S_\ell. \end{cases}$$

In all cases except for the last one, the result follows. In the last case, the assumption that there is no bad structure in $H - X$ implies that $H[S_\ell \setminus X]$ must be disconnected with u and v in separate connected components just as in the proof of Lemma 5.2. If u and v are also in separate connected components in $H[S_\ell]$, then it is easy to see that the lemma holds.

Otherwise, u and v are in the same connected component in $H[S_\ell]$; we denote its edges by E . Since u and v are in separate connected components of $H[S_\ell \setminus X]$, we know that $E \setminus \mathcal{C}(X)$ has separate connected components with edge sets E_u and E_v containing u and v , respectively. Let $S_u = \Gamma(E_u)$, $S_v = \Gamma(E_v)$, and $S_{\text{rest}} = S_\ell \setminus (S_u \cup S_v)$. Let $X_u = X \cap S_u$ and $\alpha_u = \alpha_{X_u}$. Define X_v , X_{rest} , α_v , and α_{rest} in the same way. We can then write $\Pi'_{S_\ell}(u = a \wedge v = b \wedge X_\ell = \alpha_\ell)$ as

$$\Pi'_{S_u}(u = a \wedge X_u = \alpha_u) \cdot \Pi'_{S_v}(v = b \wedge X_v = \alpha_v) \cdot \Pi'_{S_{\text{rest}}}(X_{\text{rest}} = \alpha_{\text{rest}}).$$

Since $\Pi'_{S_{\text{rest}}}(X_{\text{rest}} = \alpha_{\text{rest}})$ depends only on α , the lemma follows. \square

Using this lemma, we can prove that $M_{X,\alpha}$ is PSD when $H - X$ has high enough expansion.

Lemma 7.4. *Let $X \subseteq [n]$ such that $H - X$ is $(r, k - t/2 + \varepsilon)$ -expanding for $r = \omega(1)$ and some constant $\varepsilon > 0$. Then for any $\alpha \in \{0, 1\}^X$ with $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$, $M_{X,\alpha}$ is positive semidefinite.*

Proof. By Lemma 2.14, $M_{X,\alpha}$ is PSD if and only if $\Sigma_{X,\alpha}$ is PSD, so it suffices to show that $\Sigma_{X,\alpha}$ is PSD. The conditional distributions $\{\Pi_{S|X=\alpha}\}$ are r -locally consistent for $r = \Omega(n^{\frac{\delta}{t-2}})$ by Corollary 2.25. Then Lemma 4.2 implies that $\Sigma_{X,\alpha}$ is PSD if the correlation graph of the $\{\Pi_{S|X=\alpha}\}$ distributions has connected components of size at most r . Lemma 7.3 implies that correlations under $\{\Pi_{S|X=\alpha}\}$ induce bad structures in $H - X$, and we can apply Theorem 6.2 to $G_{\text{bad}}(H - X)$ to complete the proof. \square

Finally, we show that for any X , $M_{X,\alpha}$ can be expressed as a nonnegative combination of $M_{\text{Cl}(X),\beta}$'s for $\beta \in [q]^{\text{Cl}(X)}$. As Lemma 7.4 implies that each $M_{\text{Cl}(X),\beta}$ is PSD, $M_{X,\alpha}$ is a nonnegative combination of PSD matrices. This implies that $M_{X,\alpha}$ is PSD for any small enough X and any $\alpha \in [q]^X$ with $\mu_C(\alpha) > 0$ for all $C \in \mathcal{C}(X)$, completing the proof of Lemma 7.1.

Claim 7.5. *Assume that $\{D_S\}$ is a family of r -locally consistent distributions. Then*

$$M_{X,\alpha} = \sum_{\substack{\beta \in [q]^T \\ \beta_X = \alpha}} D_{T|X=\alpha}(T = \beta) \cdot M_{T,\beta}$$

for any $X \subseteq T$ such that $|T| \leq r$.

The proof of this claim is immediate from the definitions of $D_{T|X=\alpha}$ and $M_{X,\alpha}$.

Acknowledgments

The first-named author would like to thank Osamu Watanabe for his encouragement. The second-named author would like to thank Anupam Gupta and Ryan O'Donnell for several helpful discussions. We would also like to thank several anonymous reviewers for constructive comments.

References

- [ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 171–180, 2010.
- [AOW15] Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 689–708, 2015.
- [BCK15] Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 97–106, 2015.
- [BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8:269–289, 2012.
- [BKS13] Boaz Barak, Guy Kindler, and David Steurer. On the Optimality of Semidefinite Relaxations for Average-Case and Generalized Constraint Satisfaction. In *Proceedings of the 4th Innovations in Theoretical Computer Science conference*, 2013.
- [BM15] Boaz Barak and Ankur Moitra. Tensor Prediction, Rademacher Complexity and Random 3-XOR. *CoRR*, abs/1501.06521, 2015.
- [BOGH⁺06] Joshua Buresh-Oppenheimer, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory Comput.*, 2:65–90, 2006.
- [BSB02] Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 9(3), 2002.
- [CLP02] A Crisanti, L Leuzzi, and G Parisi. The 3-SAT problem with large number of clauses in the ∞ -replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002.
- [CLRS13] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 350–359, 2013.
- [COGL04] Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong Refutation Heuristics for Random k -SAT. In Klaus Jansen, Sanjeev Khanna, José D.P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 3122 of *Lecture Notes in Computer Science*, pages 310–321. Springer Berlin Heidelberg, 2004.
- [Das01] Sanjeeb Dash. *On the Matrix Cuts of Lovász and Schrijver and their use in Integer Programming*. PhD thesis, Rice University, 2001.
- [DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 441–448. ACM, 2014.

- [DSS15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k . In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 59–68, 2015.
- [Fei02] Uriel Feige. Relations Between Average Case Complexity and Approximation Complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002.
- [FGK05] Joel Friedman, Andreas Goerdt, and Michael Krivelevich. Recognizing more unsatisfiable random k -SAT instances efficiently. *SIAM J. Comput.*, 35(2):408–430, 2005.
- [FO04] Uriel Feige and Eran Ofek. Easily refutable subformulas of large random 3CNF formulas. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming*, volume 3142 of *Lecture Notes in Comput. Sci.*, pages 519–530. Springer, Berlin, 2004.
- [FPV15] Vitaly Feldman, Will Perkins, and Santosh Vempala. On the Complexity of Random Satisfiability Problems with Planted Solutions. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 77–86, 2015.
- [GB] Anupam Gupta and Alex Beutel. Lecture 12 - Semidefinite Duality. Notes from course “Linear and Semidefinite Programming”.
- [GHP02] Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Complexity of semi-algebraic proofs. In *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, pages 419–430, 2002.
- [Gri01] Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613 – 622, 2001.
- [KI06] Arist Kojevnikov and Dmitry Itsykson. Lower Bounds of Static Lovász-Schrijver Calculus Proofs for Tseitin Tautologies. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, 2006.
- [LRS15] James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 567–576, 2015.
- [LS91] László Lovász and Alexander Schrijver. Cones of Matrices and Set-Functions and 0-1 Optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.
- [OW14] Ryan O’Donnell and David Witmer. Goldreich’s PRG: Evidence for near-optimal polynomial stretch. In *Proceedings of the 29th Annual Conference on Computational Complexity*, pages 1–12, 2014.
- [Rag08] Prasad Raghavendra. Optimal Algorithms and Inapproximability Results for Every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 245–254, 2008.
- [SA90] Hanif Sherali and Warren Adams. A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.
- [Sch08] Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k -CSPs. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602, 2008.
- [TW13] Madhur Tulsiani and Pratik Worah. LS_+ lower bounds from pairwise independence. In *Proceedings of the 28th Annual Conference on Computational Complexity*, pages 121–132, 2013.
- [WJ08] Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, volume 1. Now Publishers Inc., Hanover, MA, USA, January 2008.

A Proofs from Section 2.4

Fact 2.21. $(s, k - d)$ -expansion implies $(s, k - 2d)$ -boundary expansion.

Proof. Let S be a set of at most s hyperedges. Each of the vertices in $\Gamma(S)$ is either a boundary vertex that appears in exactly one hyperedge or it appears in two or more hyperedges, so $|\Gamma(S)| \leq |\partial S| + \frac{1}{2}(|k|S| - |\partial S|)$. Therefore, we can write

$$|\partial S| \geq 2|\Gamma(S)| - k|S| \geq (k - 2d)|S|,$$

where the second inequality follows the expansion assumption. \square

Lemma 2.22. Fix $\delta > 0$. With high probability, a set of $m \leq n^{t/2-\varepsilon}$ constraints chosen uniformly at random is both $(n^{\frac{\varepsilon}{t-2}}, k - \frac{t}{2} + \frac{\varepsilon}{2})$ -expanding and $(n^{\frac{\varepsilon}{t-2}}, k - t + \varepsilon)$ -boundary expanding.

Proof. By Fact 2.21, it suffices to show that a random instance is $(n^{\frac{\varepsilon}{t-2}}, k - \frac{t}{2} + \frac{\varepsilon}{2})$ -expanding. We give the proof of [OW14], which is essentially the same as that of [BGMT12].

We want to upper bound the probability that any set of r hyperedges with $r \leq n^{\frac{\varepsilon}{t-2}}$ contains less than $r(k - \frac{t}{2} + \frac{\varepsilon}{2})$ vertices. Fix an r -tuple of edges T ; this is a tuple of indices in $[m]$ representing the indices of the hyperedges in T . We wish to upper bound $\Pr[|\Gamma(T)| \leq v]$; we can do this with the quantity

$$\frac{(\# \text{ sets } S \text{ of } v \text{ vertices}) \cdot (\# \text{ sets of } r \text{ edges contained in } S)}{(\# \text{ of ways of choosing } r \text{ edges})}.$$

Taking a union bound over all tuples of size r , we see that

$$\Pr[|\Gamma(S)| \leq v \forall S \text{ s.t. } |S| = r] \leq r! \binom{m}{r} \cdot \frac{\binom{n}{v} \binom{k!}{r} \binom{v}{k}}{(k! \binom{n}{k})^r}.$$

Simplifying and applying standard approximations, we get that

$$\Pr[|\Gamma(S)| \leq v \forall S \text{ s.t. } |S| = r] \leq e^{(2+k)r+v} v^{kr-v} r^{-r} n^{v-kr} m^r.$$

Set $v = \lfloor r(k - \frac{t}{2} + \frac{\varepsilon}{2}) \rfloor$ and simplify to get

$$\Pr \left[|\Gamma(S)| < r \left(k - \frac{t}{2} + \frac{\varepsilon}{2} \right) \forall S \text{ s.t. } |S| = r \right] \leq (C(k, t) \cdot m n^{-(t/2-\varepsilon/2)} r^{t/2-1-\varepsilon/2})^r$$

for some constant $C(k, t)$ depending on k and t . Then set $m = n^{t/2-\varepsilon}$ and take a union bound over all choices of r to get that

$$\begin{aligned} \Pr \left[H_{\mathcal{I}} \text{ not } \left(n^{\frac{\varepsilon}{t-2}}, k - \frac{t}{2} + \frac{\varepsilon}{2} \right) \text{-expanding} \right] &\leq \sum_{r=1}^{\lfloor n^{\varepsilon/(t-2)} \rfloor} (C(k, t) \cdot n^{-\varepsilon/2} r^{t/2-1-\varepsilon/2})^r \\ &= \sum_{r=1}^{\lfloor \log n \rfloor} (C(k, t) \cdot n^{-\varepsilon/2} r^{t/2-1-\varepsilon/2})^r + \sum_{r=\lfloor \log n \rfloor+1}^{\lfloor n^{\varepsilon/(t-2)} \rfloor} (C(k, t) \cdot n^{-\varepsilon/2} r^{t/2-1-\varepsilon/2})^r \\ &\leq 2C(k, t) \cdot n^{-\varepsilon/2} (\log n)^{t/2-\varepsilon/2} + n^{\frac{\varepsilon}{t-2}} (C(k, t) \cdot n^{-\varepsilon/2} (n^{\frac{\varepsilon}{t-2}})^{t/2-1-\varepsilon/2})^{\log n} \\ &= O(n^{-\varepsilon/3}). \end{aligned}$$

\square

B Equivalence between PSDness of the degree-2 moment matrix and the covariance matrix

Lemma B.1.

$$\begin{pmatrix} 1 & w^\top \\ w & B \end{pmatrix} \text{ is PSD} \iff B - ww^\top \text{ is PSD.}$$

Proof.

$$\begin{aligned} \begin{pmatrix} 1 & w^\top \\ w & B \end{pmatrix} \text{ is PSD} &\iff \left((v_0 \ v) \begin{pmatrix} 1 & w^\top \\ w & B \end{pmatrix} (v_0 \ v)^\top \geq 0 \ \forall v_0 \in \mathbb{R}, v \in \mathbb{R}^{nq} \right) \\ &\iff (v_0^2 + 2\langle w, v \rangle v_0 + \langle Bv, v \rangle \geq 0 \ \forall v_0 \in \mathbb{R}, v \in \mathbb{R}^{nq}) \\ &\iff ((v_0 + \langle w, v \rangle)^2 - \langle w, v \rangle^2 + \langle Bv, v \rangle \geq 0 \ \forall v_0 \in \mathbb{R}, v \in \mathbb{R}^{nq}) \\ &\iff (-\langle w, v \rangle^2 + \langle Bv, v \rangle \geq 0 \ \forall v \in \mathbb{R}^{nq}) \\ &\iff (v(B - ww^\top)v^\top \geq 0 \ \forall v \in \mathbb{R}^{nq}) \\ &\iff B - ww^\top \text{ is PSD.} \end{aligned}$$

□

Lemma 2.9. M is PSD if and only if Σ is PSD.

Proof. We rewrite M as

$$\begin{pmatrix} 1 & w^\top \\ w & B \end{pmatrix},$$

where w is a vector whose (i, a) -element is $D_{\{i\}}(x_i = a)$ for $i \in [n]$ and $a \in [q]$ and B is a matrix whose $((i, a), (j, b))$ -element is $D_{\{i, j\}}(x_i = a \wedge x_j = b)$ for $i, j \in [n]$ and $a, b \in [q]$. From Lemma B.1, we know that

$$\begin{pmatrix} 1 & w^\top \\ w & B \end{pmatrix} \text{ is PSD if and only if } B - ww^\top \text{ is PSD.}$$

Observe that $B - ww^\top$ is equal to the covariance matrix Σ .

□

C Proof of Lemma 2.12

Lemma 2.12. Let $X \subseteq [n]$ and let $\{D_S\}$ be a family of r -locally consistent distributions for sets $S \subseteq [n]$ such that $S \cap X = \emptyset$ and $|S \cup X| \leq r$. Then the family of conditional distributions $\{D_{S|X=\alpha}\}$ is $(r - |X|)$ -locally consistent for any $\alpha \in \{0, 1\}^X$ such that $D_X(X = \alpha) > 0$.

Proof. Tulsiani and Worah proved this lemma and we will use their proof [TW13]. Let $S \subseteq T$ and $|T \cup X| \leq r$. Let β be any assignment to S . Then local consistency of the $\{D_S\}$ distributions implies that $D_{S \cup X}(S = \beta \wedge X = \alpha) = D_{T \cup X}(S = \beta \wedge X = \alpha)$ and $D_{S \cup X}(X = \alpha) = D_{T \cup X}(X = \alpha)$. We therefore have that

$$\begin{aligned} D_{S|X=\alpha}(S = \beta) &= \frac{D_{S \cup X}(S = \beta \wedge X = \alpha)}{D_{S \cup X}(X = \alpha)} \\ &= \frac{D_{T \cup X}(S = \beta \wedge X = \alpha)}{D_{T \cup X}(X = \alpha)} = D_{T|X=\alpha}(S = \beta). \end{aligned}$$

□

D Proofs from Section 2.5

Lemma 2.23. If $H_{\mathcal{I}}$ is (s_1, e_1) -expanding and S is a set of variables such that $|S| < (e_1 - e_2)s_1$ for some $e_2 \in (0, e_1)$, then there exists a set $\text{Cl}(S) \subseteq [n]$ such that $S \subseteq \text{Cl}(S)$ and $H_{\mathcal{I}} - \text{Cl}(S)$ is (s_2, e_2) -expanding with $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$ and $\text{Cl}(S) \leq \frac{e_1}{e_1 - e_2}|S|$.

Proof. We compute $\text{Cl}(S)$ using the closure algorithm of [BGMT12, TW13]:

Input: An (s_1, e_1) -expanding instance \mathcal{I} , $e_2 \in (0, e_1)$, a tuple $S = (x_1, \dots, x_u) \in [n]^u$ such that $u < (e_1 - e_2)s_2$.

Output: The closure $\text{Cl}(S)$.

Set $\text{Cl}(S) \leftarrow \emptyset$ and $s_2 \leftarrow s_1$.

for $i = 1, \dots, u$

$\text{Cl}(S) \leftarrow \text{Cl}(S) \cup \{x_i\}$

if $H_{\mathcal{I}} - \text{Cl}(S)$ is not (s_2, e_2) -expanding, **then**

Find largest set of constraints N_i in $H_{\mathcal{I}} - \text{Cl}(S)$ such that $|N_i| \leq s_2$ and $|\Gamma(N_i)| \leq e_2|N_i|$.

Break ties by lexicographic order.

$\text{Cl}(S) \leftarrow \text{Cl}(S) \cup \Gamma(N_i)$

$s_2 \leftarrow s_2 - |N_i|$

return $\text{Cl}(S)$

It is clear from the statement of the algorithm that $S \subseteq \text{Cl}(S)$. We need to show that $H_{\mathcal{I}} - \text{Cl}(S)$ is (s_2, e_2) -expanding, that $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$, and that $\text{Cl}(S) \leq \frac{e_1}{e_1 - e_2}|S|$. We give the proof of [BGMT12].

1. $H_{\mathcal{I}} - \text{Cl}(S)$ is (s_2, e_2) -expanding

We will show that $H_{\mathcal{I}} - \text{Cl}(S)$ is (s_2, e_2) -expanding at every step of the algorithm. Say we are in step i and that $H_{\mathcal{I}} - (\text{Cl}(S) \cup \{x_i\})$ is not (s_2, e_2) -expanding; if $H_{\mathcal{I}} - (\text{Cl}(S) \cup \{x_i\})$ were (s_2, e_2) -expanding, we would be done. Let N_i be the largest set of hyperedges in $H_{\mathcal{I}} - \text{Cl}(S)$ such that $|N_i| \leq s_2$ and $|\Gamma(N_i)| \leq e_2|N_i|$. We need to show that $H_{\mathcal{I}} - (\text{Cl}(S) \cup \{x_i\} \cup \Gamma(N_i))$ is $(s_2 - |N_i|, e_2)$ -expanding.

To see this, assume for a contradiction that there exists a set of hyperedges N' in $H_{\mathcal{I}} - (\text{Cl}(S) \cup \{x_i\} \cup \Gamma(N_i))$ such that $N' \leq s_2 - |N_i|$ and $|\Gamma(N')| < e_2|N'|$. Consider $N_i \cup N'$. Note that N_i and N' are disjoint, so $|N_i \cup N'| \leq s_2$. Also, $|\Gamma(N_i \cup N')| \leq e_2|N_i| + e_2|N'| = e_2|N_i \cup N'|$. This contradicts the maximality of N_i .

2. $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$

Consider the set $N = \bigcup_{i=1}^u N_i$. First, note that $|N| = s_1 - s_2$, so $|\Gamma(N)| \geq e_1(s_1 - s_2)$ by expansion of $H_{\mathcal{I}}$. Second, each element of $\Gamma(N) - S$ occurs in exactly one of the N_i 's and each N_i has expansion at most e_2 . Using these two observations, we see that

$$e_1(s_1 - s_2) \leq |\Gamma(N)| \leq |S| + \sum_{i=1}^u e_2|N_i| = |S| + e_2(s_1 - s_2).$$

This implies the claim.

3. $\text{Cl}(S) \leq \frac{e_1}{e_1 - e_2}|S|$

Observe that $\text{Cl}(S) = S \cup \bigcup_{i=1}^u \Gamma(N_i)$. Also, every N_i has expansion at most e_2 . Therefore, we have that

$$\begin{aligned} |\text{Cl}(S)| &\leq |S| + \sum_{i=1}^u |\Gamma(N_i)| \\ &\leq |S| + e_2 \sum_{i=1}^u |N_i| \\ &\leq |S| + \frac{e_2 |S|}{e_1 - e_2} \\ &= \left(\frac{e_1}{e_1 - e_2} \right) |S|, \end{aligned}$$

where we used that $\sum_{i=1}^u |N_i| = s_1 - s_2$ and $s_2 \geq s_1 - \frac{|S|}{e_1 - e_2}$. \square

Theorem 2.24. For a random instance \mathcal{I} with $m \leq \Omega(n^{t/2-\varepsilon})$, the family of distributions $\{\Pi_S\}_{|S| \leq r}$ is r -locally consistent for $r = n^{\frac{\varepsilon}{t-2}}$ and is supported on satisfying assignments.

To prove the theorem, we will use the following lemma, which says that the local distributions Π'_S and Π'_T with $S \subseteq T$ are consistent if $H_{\mathcal{I}} - S$ has high boundary expansion.

Lemma D.1. Let P be a $(t-1)$ -wise uniform supporting predicate, let \mathcal{I} be an instance of $\text{CSP}(P)$, and let $S \subseteq T$ be sets of variables. If $H_{\mathcal{I}}$ and $H_{\mathcal{I}} - S$ are $(r, k - t + \varepsilon)$ -boundary expanding for some $\varepsilon > 0$ and $|\mathcal{C}(T)| \leq r$, then for any $\alpha \in [q]^S$, $\Pi'_S(S = \alpha) = \Pi'_T(S = \alpha)$.

First, we will use this lemma to prove Theorem 2.24.

Proof of Theorem 2.24. Let $S \subseteq T$ be sets of variables with $|T| \leq r$. Consider $U = \text{Cl}(S) \cup \text{Cl}(T)$. We will show that both Π_S and Π_T are consistent with U and therefore must themselves be consistent. Observe that $|\text{Cl}(S)|$ and $|\text{Cl}(T)|$ are at most $\frac{2kr}{\varepsilon}$, so $|U| \leq \frac{4kr}{\varepsilon}$. We want to apply Lemma D.1, so we will first show that $|\mathcal{C}(U)| \leq \frac{8r}{\varepsilon}$. Assume for a contradiction that C is a subset of $\mathcal{C}(U)$ of size $\frac{8r}{\varepsilon}$. Then

$$\frac{|\Gamma(C)|}{|C|} \leq \frac{|U|}{|C|} = \frac{4kr/\varepsilon}{8r/\varepsilon} = \frac{k}{2} < k - \frac{t}{2} + \frac{\varepsilon}{2},$$

which violates expansion (Lemma 2.22).

We know that $H_{\mathcal{I}} - \text{Cl}(T)$ and $H_{\mathcal{I}} - \text{Cl}(S)$ are $(r, k - t + \varepsilon)$ -boundary expanding for some $\varepsilon > 0$. We can then apply Lemma D.1 twice with sets $\text{Cl}(S) \subseteq U$ and $\text{Cl}(T) \subseteq U$ to see that

$$\Pi_S(S = \alpha) = \Pi'_{\text{Cl}(S)}(S = \alpha) = \Pi'_U(S = \alpha) = \Pi'_{\text{Cl}(T)}(S = \alpha) = \Pi_T(S = \alpha). \quad \square$$

Now we prove Lemma D.1.

Proof of Lemma D.1. We follow the proof of Benabbas et al. [BGMT12]. Let $\mathcal{C}(T) \setminus \mathcal{C}(S) = \{C_1, \dots, C_u\}$ and, for a constraint C , let $\sigma(C)$ be the variables in the support of C . First, observe that

$$\begin{aligned} Z_T \sum_{\substack{\beta \in [q]^T \\ \beta_S = \alpha}} \Pi'_T(\beta) &= \sum_{\gamma \in [q]^{T \setminus S}} \prod_{C \in \mathcal{C}(T)} \mu_C((\alpha, \gamma)) \\ &= \left(\prod_{C \in \mathcal{C}(S)} \mu_C(\alpha) \right) \sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^u \mu_{C_i}((\alpha, \gamma)) \\ &= (Z_S \Pi'_S(\alpha)) \sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^u \mu_{C_i}((\alpha, \gamma)) \end{aligned}$$

To finish the proof, we need the following claim.

Claim D.2. *There exists an ordering $(C_{i_1}, \dots, C_{i_u})$ of constraints of $\mathcal{C}(T) \setminus \mathcal{C}(S)$ and a partition V_1, \dots, V_u, V_{u+1} of variables of $T \setminus S$ such that for all $j \leq u$ the following hold.*

1. $V_j \subseteq \sigma(C_{i_j})$.
2. $|V_j| \geq k - t + 1$
3. V_j does not intersect $\sigma(C_{i_l})$ for any $l > j$. That is, $V_j \cap \bigcup_{l>j} \sigma(C_{i_l}) = \emptyset$.

Proof of Claim D.2. We will find the sets V_j by repeatedly using $(r, k - t + \varepsilon)$ -boundary expansion of $H_{\mathcal{T}} - S$. Let $Q_1 = \mathcal{C}(T) \setminus \mathcal{C}(S)$. We know that $|Q_1| \leq r$, so boundary expansion of $H_{\mathcal{T}} - S$ implies that $|\partial(Q_1) \setminus S| \geq (k - t + \varepsilon)|Q_1|$. There must exist a constraint $C_j \in Q_1$ with at least $k - t + 1$ boundary variables in $H_{\mathcal{T}} - S$; i.e., $|\sigma(C_j) \cap (\partial(Q_1) \setminus S)| \geq k - t + 1$. We then set $V_1 = \sigma(C_j) \cap (\partial(Q_1) \setminus S)$ and $i_1 = j$. Let $Q_2 = Q_1 \setminus C_j$. We apply the same process $u - 1$ more times until Q_l is empty and then set $V_{u+1} = (T \setminus S) \setminus (\bigcup_{j=1}^u V_j)$. We remove constraint C_{i_l} at every step and $V_l \subseteq \sigma(C_{i_l})$, so it holds that $V_j \cap \bigcup_{l>j} \sigma(C_{i_l}) = \emptyset$. \square

Using the claim, we can write $\sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^u \mu_{C_{i_l}}((\alpha, \gamma))$ as

$$\sum_{\gamma_{u+1} \in [q]^{V_{u+1}}} \sum_{\gamma_u \in [q]^{V_u}} \mu_{C_u}(\gamma'_u) \sum_{\gamma_{u-1} \in [q]^{V_{u-1}}} \mu_{C_{u-1}}(\gamma'_{u-1}) \cdots \sum_{\gamma_1 \in [q]^{V_1}} \mu_{C_1}(\gamma'_1),$$

where each γ'_j depends on α and γ_l with $l \geq j$ but does not depend on γ_l with $l < j$. We will evaluate this sum from right to left. We know that each V_j contains at least $k - t + 1$ elements, so $(t - 1)$ -wise uniformity of μ implies that $\sum_{\gamma_j \in [q]^{V_j}} \mu_{C_j}(\gamma'_j) = q^{-(k - |V_j|)}$. Applying this repeatedly, we see that

$$\sum_{\gamma \in [q]^{S \setminus T}} \prod_{i=1}^u \mu_{C_{i_l}}((\alpha, \gamma)) = q^{-(ku - \sum_{j=1}^{u+1} |V_j|)} = q^{|T \setminus S| - k|\mathcal{C}(T) \setminus \mathcal{C}(S)|}.$$

Plugging this quantity into the above calculation, we obtain

$$Z_T \sum_{\substack{\beta \in [q]^T \\ \beta|_S = \alpha}} \Pi'_T(\beta) = Z_S \Pi'_S(\alpha) q^{|T \setminus S| - k|\mathcal{C}(T) \setminus \mathcal{C}(S)|}.$$

Since $H_{\mathcal{T}}$ has $(r, k - t + \varepsilon)$ -boundary expansion for some $\varepsilon > 0$, we can set $S = \emptyset$ to get that $Z_T = q^{|T| - k|\mathcal{C}(T)|}$. Similarly, $Z_S = q^{|S| - k|\mathcal{C}(S)|}$. Plugging these two quantities in completes the proof. \square

E Equivalence of Sherali-Adams, SA_+ , and static LS_+ tightenings of linear and degree- k relaxations of $\text{CSP}(P)$

Lemma 2.3. Let $r \geq k$ and let \mathcal{I} be an instance of $\text{CSP}(P)$ with binary alphabet. Then the following statements hold.

1. $\text{SA}^r(R_{\mathcal{I}}) \subseteq \text{SA}^{r+k+1}(L_{\mathcal{I}})$ and $\text{SA}^r(L_{\mathcal{I}}) \subseteq \text{SA}^{r+k+1}(R_{\mathcal{I}})$.
2. $\text{SA}_+^r(R_{\mathcal{I}}) \subseteq \text{SA}_+^{r+k+1}(L_{\mathcal{I}})$ and $\text{SA}_+^r(L_{\mathcal{I}}) \subseteq \text{SA}_+^{r+k+1}(R_{\mathcal{I}})$.
3. $\text{StaticLS}_+^r(R_{\mathcal{I}}) \subseteq \text{StaticLS}_+^{r+k+1}(L_{\mathcal{I}})$ and $\text{StaticLS}_+^r(L_{\mathcal{I}}) \subseteq \text{StaticLS}_+^{r+k+1}(R_{\mathcal{I}})$.

Proof. First, we recall some notation from Section 2. Let $P'(x)$ be the unique degree- k polynomial such that $P'(z) = P(z)$ for all $z \in \{0, 1\}^k$; assume P and P' depend on all k of their input variables. Let $F = \{z \in \{0, 1\}^k \mid P(z) = 0\}$. For $b \in \{0, 1\}$, define $a^{(b)}$ so that $a^{(b)}$ is a if $b = 0$ and $1 - a$ if $b = 1$. For $z \in [0, 1]^k$

and $c \in \{0, 1\}^k$, we define $z^{(c)} \in [0, 1]^k$ so that $(z^{(c)})_i = z_i^{(c_i)}$. For $f \in \{0, 1\}^k$ and $z \in [0, 1]^k$, let $P_f(z) = \sum_{i=1}^k z^{(f_i)}$. Let $(c, S) \in \mathcal{I}$ be any constraint. Note that

$$P'(x_S^{(c)}) - 1 = \sum_{f \in F} 1_{\{x_S^{(c)}=f\}}(x) \cdot (P_f(f) - 1) = - \sum_{f \in F} 1_{\{x_S^{(c)}=f\}}(x). \quad (\text{E.7})$$

We give the proof for SA. The SA₊ and static LS₊ cases are identical, as constraints in SA₊ and static LS₊ generate exactly the same lifted constraints as in SA.

First, assume that we have a family of $(r + k + 1)$ -locally consistent distributions $\{D_S\}$ satisfying

$$\mathbb{E}_D[1_{\{x_T=\alpha\}}(x) \cdot (P_f(x_S^{(c)}) - 1)] \geq 0 \quad (\text{E.8})$$

for all $f \in F$, $T \subseteq [n]$, and $\alpha \in \{0, 1\}^{|T|}$ such that $\deg(1_{\{x_T=\alpha\}}(x) \cdot (P'_f(x_S^{(c)}) - 1)) \leq r + k + 1$. Fix any $U \subseteq [n]$ and $\beta \in \{0, 1\}^{|U|}$ such that $\deg(1_{\{x_U=\beta\}}(x) \cdot (P'(x_S^{(c)}) - 1)) \leq r$.

We want to show that $\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot (P'(x_S^{(c)}) - 1)] = 0$. Using (E.7), it suffices to show that

$$\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)] = 0$$

for all $f \in F$. First, we need to bound the size of $|U|$. Again using (E.7), we know that

$$\begin{aligned} 1_{\{x_U=\beta\}}(x) \cdot (P'(x_S^{(c)}) - 1) &= -1_{\{x_U=\beta\}}(x) \sum_{f \in F} 1_{\{x_S^{(c)}=f\}}(x) \\ &= -1_{\{x_{U \setminus S}=\beta_{U \setminus S}\}}(x) \sum_{f \in F} 1_{\{x_{S \cap U}=\beta_{S \cap U}\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x). \end{aligned}$$

We have two cases. If the assignments $x_{S \cap U} = \beta_{S \cap U}$ and $x_S^{(c)} = f$ are inconsistent for all $f \in F$, then $1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) = 0$ for all $f \in F$ after multilinearization. Then $\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)] = 0$ for all $f \in F$ and we are done.

Otherwise,

$$\sum_{f \in F} 1_{\{x_{S \cap U}=\beta_{S \cap U}\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) \neq 0$$

and $\deg(1_{\{x_U=\beta\}}(x) \cdot (P'(x_S^{(c)}) - 1)) \leq r$ implies that $\deg(1_{\{x_{U \setminus S}=\beta_{U \setminus S}\}}(x)) \leq r$. Then $|U \setminus S| \leq r$ and $|U \cup S| \leq r + k$. Since $1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) \geq 0$ for all $x \in \{0, 1\}^n$ and $1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)$ depends on at most $r + k$ variables, we know that

$$\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)] \geq 0. \quad (\text{E.9})$$

On the other hand, assumption (E.8) implies that

$$-\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)] = \mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) \cdot (P_f(x_S^{(c)}) - 1)] \geq 0$$

since $\deg(1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) \cdot (P_f(x_S^{(c)}) - 1)) \leq r + k + 1$.

For the other direction, assume that we have a family of $(r + k + 1)$ -locally consistent distributions $\{D_S\}$ satisfying

$$\mathbb{E}_D[1_{\{x_T=\alpha\}} \cdot (P'(x_S^{(c)}) - 1)] = 0 \quad (\text{E.10})$$

for all $T \subseteq [n]$ and $\alpha \in \{0, 1\}^{|T|}$ such that $\deg(1_{\{x_T=\alpha\}} \cdot (P'(x_S^{(c)}) - 1)) \leq r + k + 1$. Fix any $U \subseteq [n]$ and $\beta \in \{0, 1\}^{|U|}$ such that $\deg(1_{\{x_U=\beta\}}(x) \cdot (P_f(x_S^{(c)}) - 1)) \leq r$.

We want to show that $\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot (P'_f(x_S^{(c)}) - 1)] \geq 0$. We will do this by proving that

$$\mathbb{E}_D[1_{\{x_S^{(c)}=z\}}(x) \cdot 1_{\{x_U=\beta\}}(x) \cdot (P_f(x_S^{(c)}) - 1)] \geq 0$$

for any $z \in \{0, 1\}^k$ and then summing over all z .

If $S \subseteq U$ and β assigns S to f , then $1_{\{x_U=\beta\}}(x) \cdot (P_f(x_S^{(c)}) - 1) = 0$ after multilinearizing and we are done. Otherwise, it is easy to see that $|U| \leq r$.

We consider two cases: $z = f$ and $z \neq f$. In the first case, we can use (E.10) and (E.7) to see that

$$\sum_{f \in F} \mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x)] = 0.$$

By (E.9), each term in the sum must be 0, so we have that

$$\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=f\}}(x) \cdot (P_f(x_S^{(c)}) - 1)] = -\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=z\}}(x)] = 0.$$

When $z \neq f$, (E.9) and the fact that $P_f(z) - 1 \geq 0$ imply that

$$\mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=z\}}(x) \cdot (P_f(x_S^{(c)}) - 1)] = (P_f(z) - 1) \cdot \mathbb{E}_D[1_{\{x_U=\beta\}}(x) \cdot 1_{\{x_S^{(c)}=z\}}(x)] \geq 0. \quad \square$$

F Correspondence between static LS_+ proof system and SDP relaxation

Say we start with a set of constraints $A = \{g_1(x) \geq 0, g_2(x) \geq 0, \dots, g_m(x) \geq 0\}$. Recall that an r -round static LS_+ solution is a set of local distributions $\{D_S\}$ satisfying the following conditions.

1. $\{D_S\}_{S \subseteq [n], |S| \leq r}$ is r -locally consistent.
2. $\mathbb{E}_{x \sim D}[1_{\{x_T=\alpha\}}(x) \cdot g(x)] \geq 0$ for all $g \in A$, $T \subseteq [n]$, and $\alpha \in [q]^{|T|}$ such that $\deg(1_{\{x_T=\alpha\}}(x) \cdot f(x)) \leq r$.
- 3'. $M_{X,\alpha}$ is PSD for all $X \subseteq [n]$ with $|X| \leq r - 2$ and all $\alpha \in [q]^X$.

A static LS_+ refutation has the form

$$\sum_i \gamma_i \cdot b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x) + \sum_j (x_j^2 - x_j) h_j(x) = -1, \quad (\text{F.11})$$

where $\gamma_i \geq 0$, b_i is an axiom or the square of an affine function, and the h_j 's are arbitrary polynomials.

Proposition F.1. *The r -round static LS_+ SDP is infeasible if and only if a degree- r static LS_+ refutation exists.*

Proof. Using linearity of $\mathbb{E}[\cdot]$, we can write the first two SA conditions as linear constraints in the variables $\mathbb{E}_D[1_{\{x_S=\beta\}}(x)]$. The final constraint requires that the matrices $M_{T,\alpha}$, whose entries are also variables of the form $\mathbb{E}_D[1_{\{x_S=\beta\}}(x)]$, is PSD. As mentioned above, we can arrange the matrices $M_{T,\alpha}$ into a block diagonal matrix \mathcal{M} such that \mathcal{M} is PSD if and only if each of the $M_{T,\alpha}$'s are PSD. Let d be the dimension of \mathcal{M} . We can think of the r -round SA constraints as being linear constraints on the entries of \mathcal{M} . In particular, say these linear SA constraints have the form $A \cdot \text{vec}(\mathcal{M}) \geq b$, where $\text{vec}(\mathcal{M}) \in \mathbb{R}^{d^2}$ is the vector formed by concatenating the columns of \mathcal{M} . Let c be the number of rows of A . Then we can write the static LS_+ SDP as

$$\begin{aligned} A \cdot \text{vec}(\mathcal{M}) &\geq b \\ \mathcal{M} &\succeq 0. \end{aligned}$$

First, we show that the existence of a degree- r static LS_+ refutation implies that the r -round static LS_+ SDP is infeasible. Assume for a contradiction that there exists a family of local distributions $\{D_S\}$ satisfying the three constraints above. We will derive a contradiction by applying $E_D[\cdot]$ to each term of (F.11). Specifically, we will show that if $\deg(b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x)) \leq r$, then $\mathbb{E}_D[\gamma_i \cdot b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x)] \geq 0$. Applying $\mathbb{E}_D[\cdot]$ to the left-hand side of (F.11) gives value at least 0, applying $\mathbb{E}_D[\cdot]$ to the right-hand side gives value -1 , and we obtain a contradiction. To show that $\mathbb{E}_D[\gamma_i \cdot b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x)] \geq 0$, we will consider two cases.

Case 1: b_i is an axiom.

This case is immediate from Condition 2.

Case 2: b_i is the square of a linear form.

This case follows almost immediately from Condition 3'. Write $b_i(x)$ as follows:

$$b_i(x) = \left(a_0 + \sum_{u \in [n]} a_u x_u \right)^2 = \sum_{u,v \in [n]} a_u a_v x_u x_v + 2a_0 \sum_{u \in [n]} a_u x_u + a_0^2.$$

Then we have the following calculation:

$$\begin{aligned} \mathbb{E}_D[b_i(x) \cdot 1_{\{x_{T_i}=\alpha_i\}}(x)] &= \sum_{u,v \in [n]} a_u a_v \cdot \mathbb{E}_D[x_u x_v \cdot 1_{\{x_{T_i}=\alpha_i\}}] + 2a_0 \sum_{u \in [n]} a_u \cdot \mathbb{E}_D[x_u \cdot 1_{\{x_{T_i}=\alpha_i\}}] \\ &\quad + a_0^2 \cdot \mathbb{E}_D[1_{\{x_{T_i}=\alpha_i\}}] \\ &= \sum_{u,v \in [n]} a_u a_v \cdot M_{T_i, \alpha_i}((u, 1), (v, 1)) \\ &\quad + 2a_0 \sum_{u \in [n]} a_u \cdot M_{T_i, \alpha_i}((u, 1), 0) + a_0^2 \cdot M_{T_i, \alpha_i}(0, 0) \\ &= (a')^\top (M_{T_i, \alpha_i}) a' \quad \text{where } a'(u, 1) = a_u \text{ and } a'(u, 0) = 0 \text{ for all } u \in [n] \text{ and } a'(0) = a_0 \\ &\geq 0 \quad \text{by Condition 3'}. \end{aligned}$$

For the other direction, assume that the r -round static LS_+ SDP is infeasible; we want to prove that a refutation (F.11) exists. Assume there exists an SA pseudoexpectation satisfying Conditions 1 and 2. Otherwise, we can find an SA refutation, which is also a valid static LS_+ refutation. Then the sets $\{\mathcal{M} \in \mathbb{R}^{d \times d} : A \cdot \text{vec}(\mathcal{M}) \geq b\}$ and $\{\mathcal{M} \in \mathbb{R}^{d \times d} : \mathcal{M} \text{ is PSD}\}$ are both nonempty, but their intersection is empty. Let $A \bullet B = \sum_{ij} A_{ij} B_{ij}$. We will need the following claim.

Claim F.2. *Let $S \subseteq \mathbb{R}^{d \times d}$ be convex, closed, and bounded. Suppose that for all $X \in S$, X is not PSD. Then there exists a PSD matrix $C \in \mathbb{R}^{d \times d}$ such that $C \bullet X < 0$ for all $X \in S$.*

Proof of Claim. The claim follows from the following two results.

Theorem F.3 (Separating Hyperplane Theorem). *Let $S, T \subseteq \mathbb{R}^d$ be closed, convex sets such that $S \cap T = \emptyset$ and S is bounded. Then there exist $a \neq 0$ and b such that*

$$a^\top x > b \text{ for all } x \in S \text{ and } a^\top x \leq b \text{ for all } x \in T.$$

Lemma F.4 (e.g., [GB, Lemma 12.4]). *If $A \bullet B \geq 0$ for all PSD B , then A is PSD.*

Applied to our situation, the Separating Hyperplane Theorem says that there exists C and δ such that $C \bullet X < \delta$ for all $X \in S$ and $C \bullet X \geq \delta$ for all PSD X . We need to show that we can choose $\delta = 0$. Applying Lemma F.4 will then complete the proof.

We know $\delta \leq 0$ because the zero matrix is PSD. It remains to show that we can choose $\delta \geq 0$. Assume for a contradiction that there exists PSD X such that $C \bullet X < 0$. We can then scale X by a large enough positive constant to get a PSD matrix X' such that $C \bullet X' < \delta$, a contradiction. \square

The claim implies that there is a PSD matrix C such that the set

$$\{\mathcal{M} \in \mathbb{R}^{d \times d} : A \cdot \text{vec}(\mathcal{M}) \geq b, C \bullet \mathcal{M} \geq 0\}$$

is empty. As this set is defined by linear inequalities, we can apply Farkas' Lemma.

Theorem F.5 (Farkas' Lemma). *Let $A \in \mathbb{R}^{m \times n}$ and consider a system of linear inequalities $Ax \geq b$. Exactly one of the following is true.*

1. *There is an $x \in \mathbb{R}^n$ such that $Ax \geq b$.*
2. *There is a $y \in \mathbb{R}^m$ such that $y \geq 0$, $y^\top A = 0$, and $y^\top b > 0$.*

In particular, this implies that there exist $y \in \mathbb{R}^c$ and $z \in \mathbb{R}$ such that $y \geq 0$, $z \geq 0$, and

$$y^\top (A \cdot \text{vec}(\mathcal{M}) - b) + zC \bullet \mathcal{M} < 0 \quad (\text{F.12})$$

for all $\mathcal{M} \in \mathbb{R}^{d \times d}$. Since C is PSD, we can write its eigendecomposition $C = \sum_\ell \lambda_\ell v_\ell v_\ell^\top$ with $\lambda_\ell \geq 0$ for all ℓ . Also, recall that \mathcal{M} is block diagonal with blocks $M_{T,\alpha}$. This block structure induces a corresponding partition of $[d]$. We can write the vector $v_\ell \in \mathbb{R}^d$ as $(v_{\ell,T,\alpha})_{T,\alpha}$ using this partition. Then the second term of (F.12) is

$$\begin{aligned} zC \bullet X &= z \sum_\ell \lambda_\ell (v_\ell v_\ell^\top) \bullet \mathcal{M} \\ &= z \sum_\ell \lambda_\ell v_\ell^\top \mathcal{M} v_\ell \\ &= z \sum_\ell \lambda_\ell \sum_{\substack{|T| \leq r-2 \\ \alpha \in \{0,1\}^T}} v_{\ell,T,\alpha}^\top M_{T,\alpha} v_{\ell,T,\alpha} \\ &= z \sum_\ell \lambda_\ell \sum_{\substack{|T| \leq r-2 \\ \alpha \in \{0,1\}^T}} \sum_{i,j \in [n]} v_{\ell,T,\alpha}(i) v_{\ell,T,\alpha}(j) M_{T,\alpha}(i,j). \end{aligned}$$

Overall, we get

$$y^\top (A \cdot \text{vec}(M) - b) + z \sum_\ell \lambda_\ell \sum_{\substack{|T| \leq r-2 \\ \alpha \in \{0,1\}^T}} \sum_{i,j \in [n]} v_{\ell,T,\alpha}(i) v_{\ell,T,\alpha}(j) M_{T,\alpha}(i,j) < 0$$

for all $\mathcal{M} \in \mathbb{R}^{d \times d}$. Finally, we substitute in indicator polynomials $1_{\{x_T = \alpha\}}(x)$ for the entries of \mathcal{M} and scale appropriately to get an LS_+ refutation of the form (F.11). For the first term, since each row a_i of A each b_i correspond to an SA constraint $a_i \cdot \text{vec}(M) \geq b_i$, this substitution gives an expression of the form

$$\sum_i y_i \cdot g_i(x) \cdot 1_{\{x_{T_i} = \alpha_i\}}(x),$$

where each $g_i(x) \geq 0$ is one of our initial constraints and $y_i \geq 0$. The second term has the form

$$z \sum_\ell \lambda_\ell \sum_{\substack{|T| \leq r-2 \\ \alpha \in \{0,1\}^T}} \left(\sum_{i \in [n]} v_{\ell,T,\alpha}(i) \cdot x_i \right)^2 1_{\{x_{T_i} = \alpha_i\}}(x). \quad \square$$